THE USE OF EFFECT SIZES IN CREDIT RATING MODELS

by

HENDRIK STEFANUS STEYN

submitted in accordance with the requirements
for the degree of

MASTER OF SCIENCE

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF P NDLOVU

DECEMBER 2014

Abstract

The aim of this thesis was to investigate the use of effect sizes to report the results of statistical credit rating models in a more practical way. Rating systems in the form of statistical probability models like logistic regression models are used to forecast the behaviour of clients and guide business in rating clients as “high” or “low” risk borrowers. Therefore, model results were reported in terms of statistical significance as well as business language (practical significance), which business experts can understand and interpret. In this thesis, statistical results were expressed as effect sizes like Cohen’s d that puts the results into standardised and measurable units, which can be reported practically. These effect sizes indicated strength of correlations between variables, contribution of variables to the odds of defaulting, the overall goodness-of-fit of the models and the models’ discriminating ability between high and low risk customers.

Key Terms

Practical significance; Logistic regression; Cohen’s d; Probability of default; Effect size; Goodness-of-fit; Odds ratio; Area under the curve; Multi-collinearity; Basel II

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Professor Principle Ndlovu for the guidance, useful comments, remarks and feedback during the process of writing this thesis. Furthermore, I would like to thank my father, Professor Faans Steyn for introducing me to statistics and more specifically the topic of practical significance and effect sizes. His passion for statistics is contagious. Also, I want to thank Jacob Cohen for looking beyond statistical significance. I would like to thank my wife and sons, who is always supporting me in whatever I do. Then most importantly I want to thank the Lord for creating me and all other human beings in such a way that we can come up with beautiful and complicated things.

DEDICATION

This thesis is dedicated to my Grandfather Professor Hendrik Stephanus (Fanie) Steyn (1920-2014), who by many is considered, as the Father of Statistics in South Africa.

DECLARATION BY STUDENT

The research work illustrated in this thesis was carried out in the University of South Africa, Department of Statistics under the supervision of Prof. Principal Ndlovu.

I, **Stefan Steyn** declare that: **The use of effect sizes in credit rating models** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I have not previously submitted this work, or part of it, for examination at UNISA for another qualification or at any other higher education institution.

Declared on the (date): 29 November 2014

Signed:

Name: **Stefan Steyn**

LIST OF ABBREVIATIONS

PD – Probability of Default

APA – American Psychological Association

SARB – South African Reserve Bank

JSE – Johannesburg Stock Exchange

ROC - Receiver Operating Characteristic

TTC – Through the Cycle

MBFA - McGregor Bureau of Financial Analysis

JIBAR - Johannesburg Interbank Agreed Rate

GDP - Gross Domestic Product

VIF - Variance Inflation Factor

LRM - Logistic Regression Models (Modelling)

SE - Standard Error

CAP - Cumulative Accuracy Profile

AUC - Area under Curve

MES - Measure of Effect Size

MLE - Maximum Likelihood Estimation

OR - Odds Ratio

OOR – Overall Odds Ratio

SST - Statistical Significance Testing

AIC – Akaike Information Criterion

SC – Schwarz Criterion

H-L GOF - Hosmer and Lemeshow Goodness-of-Fit

GLMM - Generalised Linear Mixed Models

1 Introduction

For certain kinds of applied research, it is no longer considered as acceptable to only report or conclude that results of hypotheses tests were statistically significant. Statistical significance may indicate that there is a real difference between populations or a real association between variables. However, further investigations may be needed to determine what the ”size” of the difference or the “strength” of the association is. That is, effect size has to be estimated. The purpose of this thesis is to explore the use and applicability of effect sizes in credit rating models. The focus will be more specifically on the models used to rate corporate companies in the wholesale credit environment. The two most important concepts that will feature throughout this research paper are: effect size and credit rating models. It is therefore important to give thorough definitions and background of both.

1.1 Effect sizes

Merriam-Webster Dictionary (2015) defines an “effect” is a change that results when something is done or happens and “Size” is physical magnitude, extent, or bulk. Wilkinson and the American Psychological Associates (APA) Task Force on Statistical Inference (1999) report that in statistics an effect size is a measure of the strength of the relationship between two variables in a statistical population. An effect size calculated from a sample is a descriptive statistic that conveys the estimated magnitude of a relationship without making any statement about whether the apparent relationship in the sample reflects a true relationship in the population. Steyn (2009) reports that effect sizes quantify the size of the difference in for example: means, correlations, etc. and may therefore be said to be a true measure of the significance of the difference also known as “Practical Significance”. Grissom and Kim (2005) report that whereas a test of statistical significance provides the quantified strength of evidence (derived from the p-value) that a null hypothesis is wrong, an effect size measures the degree to which such a null hypothesis is wrong. In terms of using a complete population versus using a representative sample from the population, Steyn (2009) states that drawing samples from the population is no longer necessary because modern statistical software, systems and data-warehouses are powerful enough for performing statistical analysis of complete or full population data sets. In other words, by analysing the complete population (with powerful statistical tools like SAS, Matlab, SPSS, etc.), statistical significance is not applicable anymore (remember that statistical inference is used in the context of probability samples). Now, practical significance becomes very important.

1.2 Credit rating models

In order to understand the type of models and their parameters that are dealt with in this thesis, there is a need to first understand the purpose and use of these models. The International Basel Committee on Banking Supervision of 2005 has set regulations and frameworks which are used by most international regulators (including the South African Reserve Bank) as guidelines on sound governance and supervision of financial institutions and in this case more specifically banks. Banking is a commodity business and as such, for banks to earn an adequate return of equity and compete for capital along with other industries; the banks need to be highly leveraged (capitalised). From this point of view, the primary function of the bank's capital is to absorb any loss a bank may suffer. Rules and guidelines set in the Swiss town of Basel, which is reported in the Basel II Accord (2005) determine the ground rules for the way banks around the world should account for loans they give out. These rules were formulated by the Bank for International Settlements in 1988 and revised in 2005. This framework also includes important definitions and guidelines on credit risk and how it should be rated, managed and modelled.

1.2.1 Credit risk and credit ratings

Credit risk is defined as "the risk of loss following a change in the factors that drive the credit quality of an asset". According to the Basel II Accord (2005), credit risk is actually the potential that causes a borrower or counterparty not to meet its obligations in accordance with the agreed terms. The Basel II Accord (2005) has defined credit rating as a 'summary indicator' of the risk inherent in individual credit, embodying an assessment of the risk of loss due to the default (i.e. going bankrupt) of a counter party by considering relevant quantitative and qualitative information. Credit ratings, through the use of symbols as typically set by international credit rating agencies like Moody's Investor Service (2009), can be defined as an expression of the opinion about credit quality of the issuer of securities with reference to a particular instrument.

1.2.2 Credit risk model

A credit risk model is a statistical model containing the loan applicant's characteristics that is either used to calculate a score representing the applicant's probability of default or to categorise borrowers into different default risk classes. In the event of the bank using this model for predicting credit ratings, a requirement from the Basel II Accord (2005) is that the bank has to satisfy its supervisor that the model has good predictive power and that regulatory capital requirements will not be distorted as a result of its use. The variables that are input to the model must form a reasonable set of predictors. The model must be accurate on average across the range of borrowers or facilities to which the bank is exposed and there must be no known material biases. According to the Basel II Accord (2005) a further requirement if a bank employs such models in the rating process is that the bank must document their methodologies. This document must:

a) Provide a detailed outline of the theory, assumptions and/or mathematical and empirical basis of the assignment of estimates to grades, individual obligors, exposures, or pools, and the data source(s) used to estimate the model;
b) Establish a rigorous statistical process (including out-of-time and out-of-sample or training and validation performance tests) for validating the model; and
c) Indicate any circumstances under which the model does not work effectively.

1.3 Leading to the problem

Statistical modelling is widely used in the banking environment to predict future behaviour of borrowers or counterparties and their probability of paying back whatever money they borrowed from the bank. As already discussed, this process is regulated with some strict rules from Basel II (2005) as well as the local regulating authority which is the South African Reserve Bank (SARB). This obviously leads to an emphasis and non-negotiable requirements on accurate, sound and robust statistical models.

When determining whether a statistical model is sound, one may still be using very “traditional” methods to test the significance of parameters and the overall model. One may not be taking into consideration, more modern, relevant and practical considerations using more powerful tools to express an opinion on significance and soundness of the statistical model. The audience interested in the outcome and significance of such a model may not be just people with a statistical background, but also people who would want to put the results into the context of their environment and base decisions they are going to make on their understanding of the model and its significance or importance. One may not need to use sampling any more, but rather investigate the complete population and express an opinion on the strength of relationships between variables and model results rather than only the fact that the relationships are significant.

1.3.1 Current solutions

The logistic regression model which is currently used in the bank's wholesale credit department to determine the probability of default (PD) makes use of data from companies that are listed on the Johannesburg Stock Exchange (JSE). It is normally companies with a turnover in excess of R10 billion. Logistic regression is used where the predicted outcome is a probability that the counterparty or client will default. The model consists of 48 logistic regression iterations. These iterations range from 1 month before default to 48 months before default. The annualized cumulative 12 month, 24 month, 36 month and 48 month PDs are then calculated and the maximum PD from these is selected as the final PD. This PD is also called a "trough the cycle" or TTC PD as it gives the probability of default over a one year period.

Only about 600 counterparties in the wholesale sectors which include local as well as international corporate companies contribute to the data used to fit the rating model. These customers or companies are normally of a high quality in terms of credit worthiness. This means that a subset of the complete population of companies (some companies had to be excluded because of missing data or based on their listing status), is used and that very few of these have ever defaulted. The goodness of fit of the overall model is measured by how well it can discriminate or differentiate between defaults and non-defaults (Bad vs. Good). Significance is also tested for the specific parameters or variables that entered the logistic regression models. Currently the significance tests are done by means of t-tests at the 5% significance level. Correlation tests are done on the different variables to decide which input variables to use in order to eliminate the problem of multi-collinearity.

In order to assess the model's ability to differentiate between good and bad credit risk (low or high probability of default), a ranking test is performed. Siddiqi (2006) reports that this is done by performing the Receiver Operating Characteristic (ROC) analysis and using the Gini coefficients. The ROC analysis gives a measure of the discriminatory power of the ordinal ranking of the counterparties by credit riskiness and the Gini coefficients measure the statistical dispersion or spread of the default probabilities. These tests give an indication of the model's fit for purpose and in a sense the overall goodness of fit.

1.3.2 Shortcomings

Although we can say that the relationship between the continuous independent variables (that entered the Logistic Regression model) and the dichotomous dependent variable is significant by evaluating the p-value, we can't really make "sense" of the strength of relationship that exists between these variables. The results of the tests indicated by a p-value < 0.05 that the independent variable is statistically significantly related to the outcome or dependent variable do not indicate how strongly the variables are related, especially when dealing with large data sets. There may also be more meaningful ways to present results of the significance of the individual predictor variables. The overall goodness of fit test for the model may have to be further investigated in terms of statistical tests and the use of effect sizes in order to report it in measurable and standardised units.

1.4 Objective of the thesis

1.4.1 Broad objective

The broad objective of this thesis is to investigate the using of the effect sizes as statistical tests for practical significance in predictive models used in the financial credit risk environment. It is evident that the use of effect sizes in financial model building is not a topic that seems to get much attention in banks. Although effect sizes are being used extensively in the behavioural science, they are not commonly used in the financial field. In recent discussions with some of the statisticians, actuaries and quantitative analysts at local financial institutions and banks, it became evident that they were not very familiar with the technique and how to incorporate it into what is currently done. Therefore, this study may add value to the quantitative model building process followed at banks and may even have a broader scope and application in the financial field.

1.4.2 Specific objectives

The specific objectives of this thesis are to take a logistic regression model used to predict probability of default and use practical significance to perform the following:

1) Correlation tests;
2) Variable selection tests;
3) Risk measures - Estimation of risk or performing of risk tests (assess the ability to discriminate between high and low risk); and
4) Goodness of fit tests.

These tests will be performed using logistic models (48 logistic regressions), fitted to the current corporate data from the JSE. The results and test statistics will be compared to those obtained from existing tests and methods. This will enable one to make a conclusion on whether the use of effect sizes and practical significance can improve the large corporate credit rating modelling process, and the banks' overall approach to credit modelling.

1.5 Data

1.5.1 Data sources

The aim of the large corporate PD model is to predict a credit rating based on financial information that is available from a wide range of companies which can be classified as large corporate companies. These companies are listed with the Johannesburg Stock Exchange (JSE). The JSE has extensive surveillance capabilities. The data from the JSE was sourced from McGregor Bureau of Financial Analysis (MBFA), which is one of the leading and most reputable providers of company data. MBFA has its origins in the establishment of the Bureau of Financial Analysis as part of the former Institute of Business Administration at the University of Pretoria in 1965. By 1992, a comprehensive database of standardised financial statements from JSE listed companies was set up. Since then MBFA has established itself to be known as the highest quality data provider in the local market. The database at MBFA stretches back over 40 years.

The data set used for modelling consisted of monthly observations for each entity or company listed on the JSE. Companies were from the following sectors: Mining, General Retailers, Food and Beverages, Forestry, Software and Computer services, Media, Travel and Leisure, Gas and Oil Production, Telecommunication, etc. These companies typically include: Media24, Dimension Data, MTM, Anglo Gold, CTM, etc. The data set contained information ranging from financial statement information which included financial ratios and line items, to market information which included share prices and volatilities. Data fields that were investigated, used and combined for modelling of the probability of default (PD) were Return Share Price, Johannesburg Interbank Agreed Rate (JIBAR), Non-Interest Bearing Debt, Turnover, Exchange Rate, Total Assets, Cash from Operations and CPI Growth to name but a few. More detail will be provided later on, on how these fields were combined to create the actual ratios and variables that were used as input to the models. The input data spanned over a cycle of 13 years, which means that at least one economic cycle was covered. An economic cycle as defined in Investopedia (2015) is the natural fluctuation of the economy between periods of expansion or growth and contraction also known as a recession. Factors such as gross domestic product (GDP), interest rates, levels of employment and consumer spending can help to determine the various stages of such an economic cycle. According to Economists at the International Monetary Fund, a global recession would take a slowdown in global growth to 3% or less. By this measure, four periods since 1985 qualified as recession periods: 1990–1993, 1998, 2001–2002 and 2008. This means that the data from 1995 to 2008 covered three of these periods and therefore the model took into account downturn and growth periods and was representative of what happened in the actual South African economy. The characteristics of the full data set used for modelling are summarised in Table 1.1:

Table 1.1: Full data set used for modelling (January 1995 – May 2008) including the number of observations and defaults

Period Covered	Jan 1995 – May 2008 (13 Years)
Number of Observations (monthly financials)	43,758
Number of counterparties (companies)	572
Number of defaults*	90

* *The number of observations (counterparties/companies) where default was detected. This data set will be used to generate 48 sub-sets of data as will be discussed in Section 1.5.3.*

It is important to note that for this study, the names of the specific companies as listed on the JSE was masked in order not to disclose any sensitive or confidential information. The purpose of the study was more on the methods used to report statistical and practical significance than on determining the actual default rating or credit worthiness of the specific companies or counterparties. Therefore, where applicable, companies or counterparties were referred to as: Company 1, 2,...etc.

1.5.2 Data management

Before the data was used for modelling, the data needed to go through a preparation process. This process can be described as a data validation or clean-up procedure. Hair *et al*. (1998) report that data cleaning is the removal of random and systematic errors from data through filtering, merging and translation. As part of this it is important to define exclusions upfront to avoid using data that is unfit for purpose. MBFA provides data on all listed companies, but some of these companies are not large corporate companies. The original data set included financial institutions and real estate companies as well and because their exposures are different from those of large corporate companies, separate models are developed to predict their credit ratings and PDs. These models fall outside the ambit of this study. Other exclusions which are mainly because of data constraints included: newly listed companies, companies with large amounts of missing data and companies that were delisted or have already defaulted.

Descriptive statistics were used to study the distributions of the variables in the data and to identify missing observations and outliers. The companies and monthly financial observations which follow were excluded from the data set.

- Financial and Real Estate sectors;
- Venture Capital and Development Capital companies;
- Companies where financial information is not yet available;
- Penny stock where Turn Over is less than R20million;
- The first 12 months for which a company is listed; and
- Data after default date.

These exclusions are in line with the strategy and requirements of the business unit where the predictive models will be used.

1.5.3 Data set construction

For the logistic regression models to be fitted and as already discussed in Section 1.3.1, the indication of whether a company defaulted or not was the dependent variable (Y), and therefore a rigorous process needed to be followed in order to identify not only whether a company defaulted but also when it defaulted. The Basel II Accord defines default as follows:

1. When an exposure is more than 90 days in arrears (time driven element); or
2. When there is reason to believe that the exposure will not be recovered in full and the exposure is classified as such (event driven element).

The default is triggered by the earlier of the two events, 1 and 2 above. Because of the small number of defaults in the large corporate sector, additional criteria were used to flag possible defaults. This identification of defaults needs to be aligned to the portfolio’s strategy and the way that they conducted business. Certain events within companies or the status of companies at a point in time (monthly) were investigated as these are normally good indicators that event 2 as described above were triggered and if this is the case, the company was considered to have defaulted. The criteria or indicators used to flag a company as defaulted are:

- Suspended companies;
- Delisted companies;
- High risk companies, which is indicated by Rating Agency downgrades and significant changes in probability of default (a typical example would be a Moody's downgrade from a BB+ to a BB credit rating);
- Internal Defaults;
- Companies with PDs in the high probability or low credit rating buckets; and
- Companies with a steep decline in shares prices.

This investigation process where each company was measured against the above mentioned criteria resulted in identifying 90 defaults within the 572 companies included in the data set.

After validation, clean-ups and exclusions were applied, the full data set was used to construct 48 different data sets in order to create a PD term structure. The concept of a PD term structure and why for this term structure, there are 48 data sets, can be explained as follows.

One has to look at the fact that for a given rating bucket or group (say for example BB+), the default probabilities would most likely vary within a year. Why should, for example the PD of a given company in the first semester be higher or lower than the PD in the second semester of the year? Intuitively one would have thought that this may either be the consequence of an internal change in the management of the firm or the consequence of a shift in the economic environment - the economy was either in a better or worse state than in the first semester and this had an impact on the financial situation of the firm. The idea was therefore to try and capture this economic cycle effect and predict how PDs varied when the economy was doing well or during a downturn (recession) period. This means that a multi-state approach was introduced which was particularly interesting during dramatic economic changes, like the 2008-2009 financial crises for instance, which typically would cause the PDs to change. There are several articles (e.g. Bangia et al. (2000), Jones (2005)) which give evidence to the relevance of this approach for the aim of having a term structure for the PDs. In those articles it is assumed that the rating migration process is a homogeneous Markov chain. On a high level, the process followed is one where models incorporate multiple equations for forecasting default at different forward time intervals, conditional on survival to that point in time. In this case the condition on survival refers to a company which survived for month 1 but defaulted in month 2, or survived until month 2, but defaulted in month 3, up to survival in month 47 but defaulted in month 48. This resulted in 48 data sets and 48 models. The different model equations share the same inputs but they have different weightings depending on the time horizon. The current and forward conditional default probabilities were then combined to derive a full default term structure. Further details of these probability chains and annual PDs will be provided and discussed in Section 2.2.2.

These data sets were used in the 48 iterations. The iterations of the model as already explained above, consisted of 1 Month before default up to 48 Months before default data sets. There are two kinds of treatments that were applied to the 48 data sets.

1.5.4 Treatment in the data set for companies that defaulted

The 1 month before default data set consists of all observations in the database up to one month before default was observed for the specific company. Similarly, the 48 months before default data set consists of all observations in the database up to 48 months before the default was observed. This data set included all defaults that have observations of monthly financial ratios, 48 months before default. Note that no data for a defaulted counterparty was included after the default has occurred (see data exclusions). In each data set the last available observation of all the companies that defaulted was flagged as a default (1) in the logistic regression.

1.5.5 Treatment in the data set for companies that did not default

For companies that did not default, all available observations of monthly financial ratios of those companies were included in the data set and their default indicator is set to (0). The default and non-default observations were merged to form the final 1 month before default data set up to the 48 months before default data sets and a summary of 5 of the 48 data sets can be seen in Table 1.2. In Chapter 3 a snapshot of the complete data set is displayed in order to give the reader an idea of what the data looks like. This snapshot can be found in Tables 3.1.1 and Table 3.1.2.

Table 1.2: A summary of 5 of the 48 data sets that were created including the number of observations and defaults for each data set.

	1 Month before Default	12 Months before Default	24 Months before Default	36 Months before Default	48 Months before Default
Companies	572	517	474	424	378
Observations (Monthly Financial Data)	43758	37743	31783	26360	21520
Defaults	90	79	68	58	44

After generating the data sets, the independent variables within these data sets were analysed and investigated (see Chapter 3) in order to come up with the optimal set of variables to use for modelling and predicting default.

1.6 Organisation of the thesis

In Chapter 2 various statistical techniques that are used to select variables and construct probability of default models were reviewed. The review includes a discussion of statistical significance as well as effect sizes for measures such as correlation, validity, odds ratios and goodness-of-fit tests. The statistical analysis of the data (PD modelling etc.) are performed in Chapter 3. The results of these analyses are reported in terms of statistical significance and also in terms of effect sizes. The conclusions using effect sizes are interpreted and compared to the conclusions using statistical significance. Chapter 4 contains the conclusion and recommendations of the thesis. Among other things, the chapter discusses how effect sizes can assist in reporting statistical results in a practical way, which will enable non-statisticians to make informed decisions based on the results of the statistical models used to predict default and the risk rating of counterparties.

2 Statistical methods of modelling probability of default

2.1 Logistic modelling of probability of default

Logistic regression modelling (LRM) is a widely used statistical method where the dependent or outcome variable (Y) is a binary or dichotomous variable (taking values of 0 for non-default and 1 for default in this thesis) and the k explanatory variables ($X_1, X_2, \ldots, X_k$) are uncorrelated continuous variables and/or 0-1 dummy variables representing the categories/levels of categorical variables/factors. In the context of this thesis the variables may include characteristics, demographic and financial information about the counterparties as well as market, economic and political indexes.

2.1.1 Logistic regression model and parameter estimation

Let Y be a dichotomous variable which is defined as

$$Y = \begin{cases} 1 & \textit{for default} \\ 0 & \textit{for non-default} \end{cases}$$

and $\mathbf{X}=(X_1, X_2, \ldots, X_k)$. Then the logistic regression model is

$$p = \Pr(Y = 1 \mid \mathbf{X}) = \frac{1}{1 + \exp[-(\alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)]}.$$

or

$$logit(p) = ln\left(\frac{p}{1-p}\right) = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \ \in (-\infty, \infty)$$

(Peng, Lee and Ingersoll (2002)), where $\alpha_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients to be estimated from the data by the maximum likelihood method and $ln\left(\frac{p}{1-p}\right)$ is called the logit-link function. It should be noted that for some data:

- the logit-link function for p may not be appropriate (e.g. the probit link function may fit the data better); and
- whatever the appropriate link function is, it may not be linearly related to all the explanatory variables (i.e.. the relationship maybe nonlinear).

The quantity $p/(1-p)$ is called the odds ratio of the event Y=1. In other words, the relationship of a dichotomous variable with its predictors is quantified with the odds ratio, which can be defined as the ratio of the odds of a "default" event occurring to the odds of a "non-default" event occurring. The e^{β_j} (j=1,2,…,k) is the odds ratio of the variable X_j for the dependent variable Y, and β_j is the change of the odds ratio of the event Y=1 per unit change in the value of predictor variable X_j (j = 1,..,k). This is well defined for continuous X_j and for X_j 0-1 dummy variables representing categories/levels of categorical variables/factors. The analysis to be performed in Chapter 3 (Section 3.4.3) of this study is limited to continuous variables.

Mays (2001) reports that logistic regression as a multivariate statistical technique is able to isolate the effect of each of the k explanatory variables ($X_1, X_2,…,X_k$) on the odds of the event occurring (default in this case) while controlling for other variables that affect the likelihood of the event. Therefore this odds ratio is referred to as the "adjusted" odds ratio because it is adjusted for the effects of the other X's. The adjusted odds ratio will be used in the analysis done in Chapter 3.

2.1.1.1 Maximum likelihood estimation of the regression coefficients

Whitehead (2011) reports that the regression coefficients α_0, β_1, β_2,.....,β_k of the model can be estimated using the method of maximum likelihood estimation (MLE). Scott Long (1997) defines the likelihood function of the regression coefficients as follows:

For i=1,2,...,n, let $\mathbf{X}_i$ =($X_{i1}, X_{i2},…,X_{ik}$) be value of the vector of explanatory variables $\mathbf{X}$=($X_1, X_2,…,X_k$) associated with Y_i representing the "0" or "1" observation from the i^{th} company, $\mathbf{X_d}$ be an n x k "design" matrix whose i^{th} row is $\mathbf{X}_i$,

$$\pi(\mathbf{X}_i) = \begin{cases} [\mathrm{P_r}(Y_i = y_i \mid \mathbf{X}_i)]^{y_i}\,[1-\mathrm{P_r}(Y_i = y_i \mid \mathbf{X}_i)]^{1-y_i},\ y_i = 0 \text{ (no-default)}, 1 \text{ (default)}; \\ \qquad\qquad\qquad\qquad\qquad\qquad 0, \text{ otherwise;} \end{cases}$$

and $\mathbf{Y}$=($Y_1,Y_2,...,Y_n$) be a vector of the independent observations. Then the likelihood function (L) of the vector of the regression coefficients $\boldsymbol{\beta}$= (α_0, β_1, β_2,.....,β_k) is given by

$$\mathrm{L}(\boldsymbol{\beta} \mid \mathbf{Y},\mathbf{X_d}) = \prod_{i=1}^{n} \pi(\mathbf{X}_i) = \prod_{i=1}^{n} [\mathrm{P_r}(Y_i = y_i \mid \mathbf{X}_i)]^{y_i}\,[1-\mathrm{P_r}(Y_i = y_i \mid \mathbf{X}_i)]^{1-y_i}.$$

The above is a function of $\boldsymbol{\beta}$ because

$$\mathrm{P_r}(Y_i = 1 \mid \mathbf{X}_i) = \frac{1}{1+\exp[-(\alpha_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k)]}.$$

The log likelihood function (LL) of **β** is:

$$\ln L(\boldsymbol{\beta} \mid \mathbf{Y},\mathbf{X_d}) = \sum_{i=1} y_i \Pr(Y_i = y_i \mid \mathbf{X}_i) + \sum_{i=1} (1-y_i)[1-\Pr(Y_i = y_i \mid \mathbf{X}_i)].$$

The maximum likelihood estimate (MLE) of **β** is the value of **β** ($\widehat{\boldsymbol{\beta}} = (\hat{\alpha}_0, \hat{\beta}_1, \hat{\beta}_2 \ldots, \hat{\beta}_k)$) which maximizes the likelihood function L or the log likelihood function (LL). The estimating equations obtained from differentiating either L or LL with respect to **β** and equating the derivatives to zero are nonlinear hence are solved using numerical methods.

The expected value of minus the second derivative of the log likelihood function (LL) with respect to **β** gives the information matrix **I(β)** about **β** in the sample, and the inverse of **I(β)** is the asymptotic variance covariance matrix of $\widehat{\boldsymbol{\beta}}$ (**V**(**β**)) which depends on **β** (Rodriguez (2001)). An estimate of **V**(**β**) is given by $\mathbf{V}(\widehat{\boldsymbol{\beta}})$.

2.1.1.2 Inference about $\boldsymbol{\beta}$ and the β_i's

When the model assumptions hold (uncorrelated explanatory variables, independent responses, appropriate link function, linearity of the relationship of the logit(p) with the explanatory variables) then the test about the model parameter vector **β** and about the individual components of **β** is the Wald test (the other asymptotically equivalent tests are the likelihood ratio and the score test) (Parzen, (1999) and Neter et al., (1996)). For testing

$H_0 : \ \boldsymbol{\beta} = \boldsymbol{\beta_0}$ (specified, e.g. as **0**) versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta_0}$

the Wald test statistic is

$$W = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})\mathbf{I}(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta_0})^{\boldsymbol{T}} \sim \chi_k^2$$

if H_0 is true. Hence H_0 is rejected at the α level of significance if W$> \chi^2_{k,(1-\alpha)}$ – the $(100 \times (1-\alpha))^{\text{th}}$ percentile of the χ_k^2 distribution or if the corresponding $p - value < \alpha$. If $H_0 : \ \boldsymbol{\beta} = \mathbf{0}$ is not rejected this means some of the *k* variables are significant in the model.

The Wald test statistic for testing hypotheses about the individual components of **β,** say

$H_0 : \ \beta_{\boldsymbol{j}} = \beta_j^0$ (specified, e.g. as 0) versus $H_1 : \beta_j \neq \beta_j^0$**,**

is

$W = \frac{\hat{\beta}_j - \beta_j^0}{se(\hat{\beta}_j)} \sim N(0,1)$ (or equivalently $W^2 \sim \chi_1^2$, where $se(\hat{\beta}_j)$ is the square root of the (j,j) element of $\mathbf{V}(\hat{\boldsymbol{\beta}})$)

if H_0 is true. Hence H_0 is rejected at the α level of significance if $W^2 > \chi^2_{1,(1-\alpha)}$ – the $(100 \times (1-\alpha))^{\text{th}}$ percentile of the χ_1^2 distribution or if the corresponding $p - value < \alpha$. If any parameter other than α_0 is not significantly different from zero, then the corresponding odds ratio (also an effect size) is not significantly different from 1, which means the corresponding explanatory variable is insignificant in the model.

For *i=1,2,…,n*, consider the following estimated probabilities of default by the *n* companies:

$$\hat{p}_i = \frac{1}{1+\exp[-(\hat{\alpha}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \ldots + \hat{\beta}_k X_{ik})]}.$$

- All companies flagged as 1 (defaulted) will have $\hat{p}_i$'s close to 1. Even non-defaulting companies but with *many* similar $\mathbf{X_i}$ characteristics to defaulting companies will have $\hat{p}_i$'s close to 1.
- All companies flagged as 0 (did not default) will have $\hat{p}_i$'s close to 0.
- The estimated coefficients are weights of the explanatory variables in the $\hat{p}_i$'s such that more weight will automatically be given to more significant variables in the $\hat{p}_i$'s.

In terms of the ranking accuracy between good and bad counterparties, as will be discussed in Section 2.1.4, if a logistic regression model fits the data well, it will be able to distinguish between good and bad (based on the Gini-coefficients). The higher default scores will be assigned to the defaulting companies in the sample, and the lower scores will be are assigned to the non-defaulting companies in the sample.

2.1.2 Model diagnostics

2.1.2.1 Goodness-of-fit tests and measures

Statistical as well as non-statistical techniques (Section 3.2.5) can be used to select the most significant and/or sensible set of explanatory variables for the fitted logistic regression model. The Wald tests (and the related asymptotically equivalent tests) discussed above this section are formal statistical methods of testing the joint significance of all or sets of explanatory variables. That is, of formally testing the goodness-of-fit of the model. The tests are to remove insignificant explanatory variables from the model. The final fitted model should then, with a desired and predetermined accuracy, predict the probability of default for any large corporate listed company. The model must exhibit a high goodness-of-fit and perform well when tested by cross-validation (also referred to as back-testing) whereby the predicted values are compared with the actual values. Hair *et al.* (1998) report that once the researcher has established that there are no violations of the underlying model assumptions (e.g. linearity, independence of the observations, etc), the next step is to assess the overall model fit with one or more goodness-of-fit measures. In standard regression, the R^2-value gives one an idea of how powerful one's model is at predicting the variable of interest. However, there is no direct equivalent of R^2 for the logistic regression model. Menard (2002) reports that some of the R^2-like (Pseudo R^2) measures which follow, are used in logistic regression modelling (LRM).

- Cox and Snell's R^2

  Cox and Snell (1989) report that this R^2 is based on the log likelihood for the model with the independent variables compared to the log likelihood for a baseline or empty model. One can also say that this R^2 is based on calculating the proportion of unexplained variance that is reduced by adding variables to the model. The problem with Cox and Snell's R^2, is that its maximum is less than 1.0, making it difficult to interpret

- Hosmer and Lemeshow's Measure

  Engelmann and Rauhmeier (2006) report that this measures how well a logit model predicts the actual probabilities of the response variable (in this case probability of default). The observations are grouped (normally 10 groups) based on percentiles of the estimated probabilities and the Hosmer and Lemenshow measure would then tell one how overall, the average expected probability per group would fit with the observed experience per that group. Hosmer and Lemeshow (2000) report that, as part of goodness-of fit testing, when fitting a logistic regression model to the data, the Hosmer and Lemeshow test calculates a probability (p) value from the chi-square distribution where the distribution have G-2 degrees of freedom and where G is the number of groups based on the percentiles of the estimated probabilities. In this case and as mentioned above, G= 10 and therefore the chi-square distribution has 8 degrees of freedom. The Hosmer and Lemeshow Goodness-of-Fit Test (H-L GOF) tests the hypotheses:

- H_0: the model is a good fit, vs.
- H_a: the model is not a good fit

If, by means of the test, one fails to reject the null hypothesis (p>0.05), it means that the model is a good fit. This process is discussed in more detail in Chapter 3 (Section 3.1.7.1)

- Nagelkerke's R^2

Nagelkerke (1991) reports that this R^2 is an adjusted version of the Cox & Snell R^2 so that the range of the R^2 is 0 to 1.This is achieved by dividing the Cox and Snell's R^2 by its maximum possible value.

These Pseudo R^2 measures will tend to be lower than traditional ordinary least squares R^2 measures. The measures and their interpretation will be discussed in more detail in Section 3.1.7.1.

Allen & Le (2008) reports that users of logistic regression models often need to assess the overall predictive strength of the fitted models, or effect sizes of the model's predictor variables. Analogues of the R^2-value have been developed, but none of these measures are interpretable on the same scale as effects of individual predictor variables. The authors further propose the overall odds ratio (which is, according to Goodyear (2011), a product of the odds ratios of the individual predictor variables in the model ($\prod_{i=1}^{k} e^{\beta_i}$) as a measure of overall effect size that is interpretable on the same scale as effects of individual predictors. For the dichotomous outcomes, the overall odds ratio (OOR) is a measure of the overall effect size for logistic regression models. The OOR is the odds ratio associated with a one-standard deviation increase in the weighted sum of the model predictor variables, where the weights are determined by each predictor variable's relative importance (which is determined by ranking their individual odds ratios). The overall odds ratio as a measure of goodness-of fit will be demonstrated in more detail in Chapter 3 (Section 3.4.4.1).

2.1.2.2 Effects of the violation of the model assumptions

Assuming that the logistic link function is appropriate for the data, then for the inferences made using the fitted model to be valid, the other model assumptions need to be met by the data. Field (2009) reports that the three assumptions are linearity of the relationship between the logit(p) and the explanatory variables, independent responses/observations, and uncorrelated explanatory variables.

a) Linearity: Hair *et al*. (1998) report that an implicit assumption of all multivariate techniques based on correlation measures of association, including LRM, is linearity. Because the Pearson's correlation coefficient only measures the strength of linear relationship between variables, one is not able to interpret or measure the strength of non-linear relationships between variables using the Spearman's correlation coefficient. This results in an underestimation of actual strength of the relationship between the dependent and the independent (predictor) variables. It is always therefore crucial to examine all relationships in order to identify any shifts from linearity. One of the simple remedies of the violation of the linearity assumption may be to transform predictor variables or create additional variables to represent the non-linear components. Jacoby (2009) reports that a method to test the assumption of linearity in the logit(p) is to use the Box-Tidwell transformation. This approach involves including a term of the form Xln(X) in the fitted model for each X, where X is a continuous predictor variable. If the coefficient for this variable is statistically significant, there is evidence of nonlinearity in the relationship between logit(p) and X.

b) Independence of the responses/observations: The assumption is that the errors in the responses should not be serially correlated if measured or observed over time. Pindyck and Rubinfeld (1991) report that a simple statistical test for serial correlation among the errors is the Durbin-Watson statistic, which is used to test for the presence of serial correlation among the residuals.

c) Multi-collinearity: Hair *et al.* (1998) report that in linear and logistic regression models, dependencies among the explanatory variables cause parameter estimates to be unstable. That is, the variances of the estimated parameters become inflated and as a consequence inferences about the parameters become incorrect. The methods of checking the presence/absence of multi-collinearity and the remedial measures are discussed in Section 2.1.3.

2.1.3 Multi-collinearity and variable selection methods

2.1.3.1 Multi-collinearity

The problem of multi-collinearity arises when some independent variables in a linear model are mutually highly correlated. This causes the model parameter estimates to be unstable (as was mentioned in Section 2.1.2.2), hence it becomes difficulty to determine the significance of the effects of the individual independent variables on the dependent variable. The severity of the problem multi-collinearity can be detected using variance inflation factors (VIF) and tolerance levels (Hair *et al.* 1998). The authors report that a tolerance of less than 0.10 and/or a VIF of 10 and higher for an independent variable indicate that the multi-collinearity problem exists.

To resolve or reduce the multi-collinearity problem, once detected, cluster analysis on variables can be used. The method groups the variables into clusters/groups in such a way that the variables within clusters are highly correlated, and variables between clusters are weakly correlated. Then a "representative'' variable is selected from each cluster in order to reduce the number of variables, as well as solve the problem of multi-collinearity. Within each cluster, the variable with a small $1\text{-}R^2$ value is the one selected, where the R^2 is the proportion of variation explained by a particular clustering of the observations.

Again, the VIF and tolerance checks for the problem multi-collinearity and cluster analysis should not be performed and interpreted in isolation from other variable selection methods discussed in Section 2.1.3.2, but should merely act as a process of the reduction or simplification process in order to determine which possible independent variables to include in the model.

2.1.3.2 Variable selection methods

Variable selection using cluster analysis based on correlations among the independent variable was discussed in Section 2.1.3.1. The question with a subjective answer is "How high is a high correlation and how low is a low correlation is?" The guideline provided by Cohen (1988) is that a correlation of 0.5 is large, a correlation of 0.3 is moderate, a correlation of 0.1 is small and anything smaller is insubstantial, trivial, or otherwise not even worth worrying about. From a business point of view cut-off points for the correlation values should be predefined. More detail about the correlation as a measure of effect size is discussed in Section 3.1.2.

There are formal statistical methods which follow the selecting the most predictive or optimal set of independent variables (Hair *et al.* 1998). In the methods the Wald or the t tests (see Section 2.1.1) are used to test the significance of the independent variables.

1. Backward selection: All the variables are included in the model, and then those that are insignificant at specified level of significance are removed from the model one-by-one (starting with the “most” insignificant variable) until only significant variables are left in the model.
2. Forward selection: The process is begun with no variables in the model. Then only significant variables are included in the model one-by-one (starting with the “most” significant) until only non-significant variables are left out of the model.
3. Stepwise selection: The process is a combination of the backward and forward selection processes in the sense that variables in the model are removed if they become insignificant upon including a significant variable during the forward selection step.

2.1.4 The validation of the model

Assessing the predictive ability of a fitted model using the data used to fit the model can obtain *biased* conclusions. There are basically two ways of dealing with this problem. The first is to split the available data randomly (or in this case it can be according to time-period) into the training (development) and validation data sets as described by Siddiqi (2006). The model is fitted by using the training data set. Then the performance of the fitted model is assessed by using the validation data set. This is done by comparing the predicted and the observed values of the dependent variable in the validation data set to those in the training data set in order to validate that there were no significant differences in the model’s predictions. However, when the amount of data available is small, this may result in an unacceptably small training data set (as well as a small validation data set) that is not representative of the population. Besides, the conclusions of the Wald tests described in Section 2.1.1 are reliable only if size of the data for fitting the models is large. In the case of a small training data set, the limitations can be overcome with resampling methods as reported by Gutierrez-Osuna (2011). This includes cross-validation, which can be performed by means of random subsampling, K-fold cross-validation and leave-one-out cross-validation or by means of bootstrapping, which is a resampling technique with replacement. These resampling methods provide an *unbiased* assessment of the model without reducing the size of the training data set.

In this thesis, the first method of validity assessment is used and therefore no further detail on the resampling methods is discussed. The training data set consists of 80-90% of all the data (which is a rule of thumb suggested by Siddiqi (2006) when working with large data sets) and the remainder is the validation data set. Once these data sets have been split, a comparison of the ability of the models fitted to these data sets to distinguish between good customers/counterparties (non-default) and bad customers/counterparties (default) is done. The model of the training set fitted to the validation data set should have similar statistical results (Gini-coefficients, R^2-values, etc.) as the model fitted to the training data set; otherwise, as reported in SAS (2011), the problems which follow may exist.

- The model is over-fitted.
- Outliers may have too much influence.
- Not modelling predictive variable(s).
- There are too many parameters in the model relative to the number of observations (over-fitting).

All of the abovementioned problems call for further investigations of the data (presence of outliers and/or influential observations) and the optimality of the selected explanatory variables into grouping companies into good and bad customer groups. In some cases if these problems exist, model redevelopment may have to be considered.

The Gini-score mentioned above and mathematically defined below, is as a measure of the ability of the model to distinguish between good counterparties and bad customers/counterparties (ranking the counterparties). Alternatively, it is a measure of the ability of the model to rank counterparties in terms of how likely they are to default. Sobehart, Keenan and Stein (2000) report that when considering a credit rating model for default rating, a high ranking score (high Gini-score) is usually an indicator of a low probability of default. In order to obtain a Cumulative Accuracy Profile (CAP) curve (sometimes also known as the Lorenz curve) from which a Gini-score is derived, one first rank companies by their default probabilities as predicted by the model, from highest to lowest. Then, out of those companies with a score higher than a value such that altogether they represent X% of the total number of companies, one records the corresponding number of defaulted companies being captured as a percentage Y% of total number of defaulted companies. Sobehart, Keenan and Stein (2000) further report that a graph of the percentage of the defaulted companies being captured against the percentage of total number of companies in the population can then be plotted (see Figure 2.1 below). Figure 2.1 displays three results from fitting three models which follow.

1. A random model, with no discriminative power between good and bad credit ratings.
2. A perfect rating model which assigns the lowest score to defaulters and which can discriminate perfectly between good and bad credit ratings for counterparties.
3. The actual or real rating model, which lies somewhere in between the above two extreme models. The performance of this actual rating model is measured by the total area under the curve (AUC).

The AUC can be used to derive the Gini-score, which is discussed in more detail below.

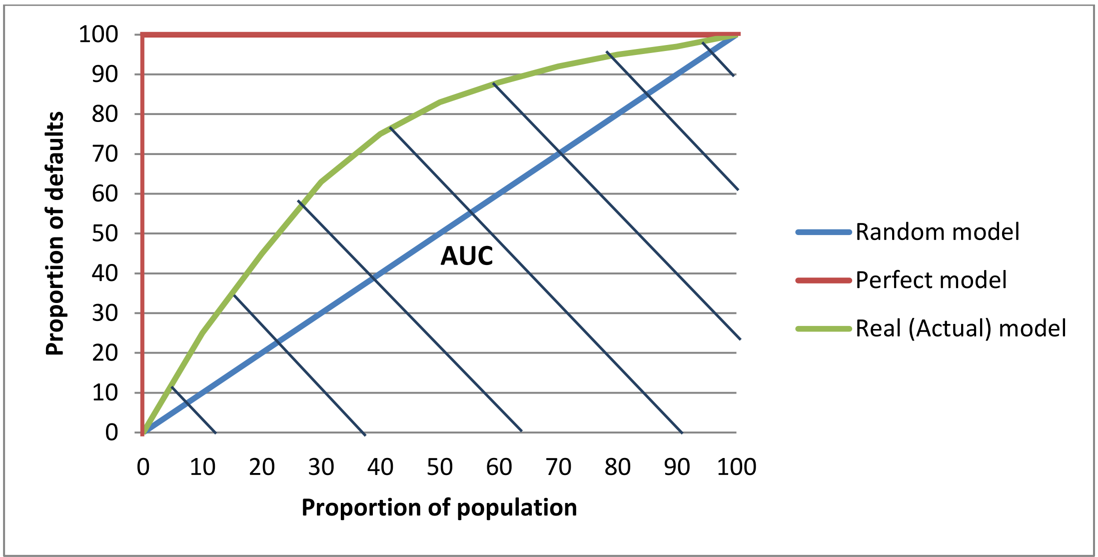


Figure 2.1: The CAP curve of the percentage of the defaulted companies predicted by the perfect, real and random models versus the percentage of total number of companies in the population (Siddiqi (2006)).

It is required to have a single measure that summarises the predictive accuracy of a model. Sobehart, Keenan and Stein (2000) further report that to calculate this single measure or score, focus should be on the area that lies above the random CAP curve and below the model (real) CAP curve (AUC) as displayed in Figure 2.1. The larger the area between the model (real) CAP curve and the random CAP curve, the better the model is in distinguishing between good and bad companies/counterparties. The ratio of the area between a real CAP curve and the random CAP curve to the area between the perfect CAP curve and the random CAP curve summarises the predictive power of the model over the entire range of possible risk ratings. This ratio measure, which lies between 0 and 1 inclusive, is called the Accuracy

Ratio or Gini-score. For example, if the area between the random CAP curve and real CAP curve is A, and the area between the real CAP curve and the perfect CAP curve is B (see Figure 2.2), then the Gini coefficient is A/(A+B). Since A+B = 0.5, then the Gini coefficient, G = 2A = 1-2B.

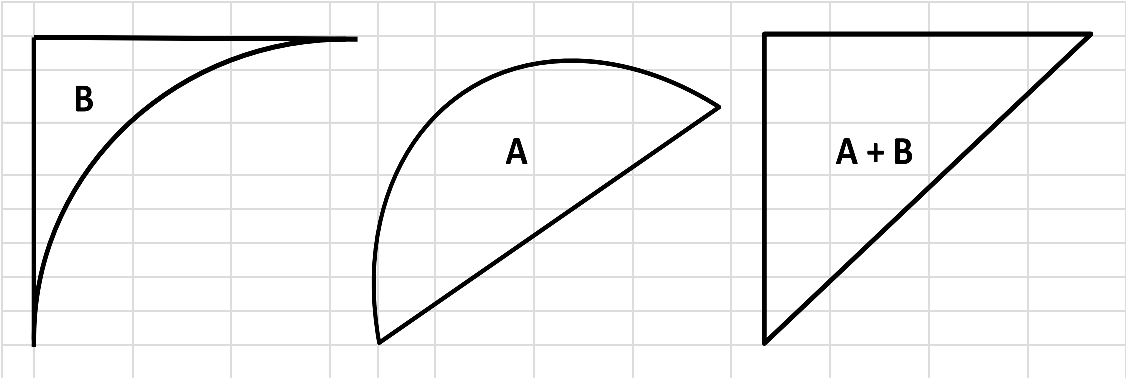


Figure 2.2: The different areas made up by the random, perfect and real CAP curves.

Satchell and Xai (2006) reported that the Gini-score can be interpreted as follows. A Gini-score of 0 means the model has no discriminative power. This comes from the fact that if the CAP curve lies on the random line resulting in an area of B = ½ and then G = 1 – 2B = 0. A score of 1 means the model can perfectly discriminate between good and bad ratings. This comes from the fact that if the CAP curve now lies on the perfect line resulting in an area of B = 0 and then G = 1 – 2B = 1. In summary, Models with Gini-scores close to 0 display little advantage over a random assignment of risk ratings, while those with Gini-scores close to 1 display almost perfect predictive power.

As mentioned earlier in this section, the abilities of the Gini-scores from the training and validation samples to discriminate between good and bad counterparties or rank the risk ratings can be compared and can therefore be used to determine the validity of the model. For a valid model, the absolute difference between the scores must be small where the rule of thumb for credit rating models is that a difference of less than 5% is considered to be small and therefore differences are insignificant.

To bring the Gini-score into the context of effect size, Rice and Harris (2005) report that effect size can be reported in terms of the AUC or receiver operating characteristic area (ROC). Rice and Harris (2005) further report that the AUC of the ROC is also a measure of the discriminatory power of the model through ordinal ranking of the counterparties according to credit riskiness. According to Siddiqi (2006), the ROC curve is a graphical plot of the (sensitivity) vs. (1 - specificity) of a binary classifier system (default or non-default) as its discrimination threshold is varied. In the context of this thesis, “sensitivity” is defined as the ratio of the “model predicted number of defaulted counterparties” to the “actual total number of defaulted counterparties”, and “specificity” is defined as “model predicted number of non-defaulted counterparties” to the “actual total number of non-defaulted counterparties”. There is a linear relationship between the ROC and Gini-score given by Flach (2009) as:

$$ROC = \frac{(\text{Gini}) + 1}{2}.$$

The AUC of the ROC together with the correlation coefficient *r* and the Cohen’s *d* (which is explained in more detail in Section 2.3.3), are Measures of Effect Size (MES) used to quantify the predictive accuracy of the fitted model. In other words, these measures can be used to check whether the current model is a valid model by comparing of the Gini-scores of the model fitted to the validation sample with the model fitted to the training sample.

2.1.5 Sensibility of choice of explanatory variables

Sometimes not all the explanatory variables included in the final model make perfect statistical sense. One would expect the modeller or model development team to be statistically strong, but at the same time have a good comprehension of the environment in which the model is to operate. There’s also need to make a qualitative assessment of the independent variables in terms of whether or not, it makes business sense to include such variables in the model. This is called business “intuition”. The final set of selected variables in the fitted model can therefore be considered sensible as long as the overall goodness of fit of the model is not compromised and a balance has been kept between quantitative modelling techniques and so called “qualitative overlays”.

2.2 The final probability of default

The construction of the 48 data sets (1 month before default to 48 months before default) and the reason for following the iterative process were discussed in Chapter 1 (Section 1.5.3). Individual modelling of these 48 data sets results in 48 fitted logistic models which when combined by using the term structure process, will produce the final probability of default (PD) model. Although for the purpose of this study, there will not be too much further investigation into the PD term structure. It is however important to understand why at a certain point, there are 48 fitted models and how to get from there to one final model which can be used to predict the PD.

The monthly default probabilities obtained from each one of the 48 fitted logistic regression models are used to construct the overall cumulative default probability term structure. The process followed to combine these probabilities is as follows. For $i=1,2,\ldots,48$, let PD_i be the probability of default of obtained from the fitted i^{th} month before default logistic model. (Equivalently, (1- PD_i) is the probability of non-default of obtained from the fitted i^{th} month before default logistic model.) Then for $m=2,3,\ldots,48$, the cumulative probability of default for month *m* is given by (Mulder, 2008):

$$CPD_m = PD_1 + \sum_{i=2}^{m} PD_i \prod_{j=1}^{i-1} (1 - PD_j) .$$

Firstly the probability from the one month before default model (PD_1) is used as is, but from month 2 onwards to month 48 the "non-default" probabilities (1 - *PD*) of each previous month is taken into account (i.e. the probability of a counterparty not defaulting in the previous month). For example, to determine the cumulative default probability for month 4, the survival probabilities for months 1 to 3 need to be considered. In other words an entity can only default in month 4 if the entity did not already default in months 1 to 3 which can be written as:

$$CPD_4 = (PD_1) + (1 - PD_1)(PD_2) + \ldots + (1 - PD_1)(1 - PD_2).(1 - PD_3)(PD_4) .$$

The cumulative default probabilities (CPD_ms) for months 12, 24, 36 and 48 (years 1, 2, 3 and 4) are annualized with the following formula (Mulder, 2008):

$$AnnPD_N = 1 - (1 - CPD_N)^{\frac{12}{N}}, N = 12, 24, 36, 48$$

The final long-run average one year probability of default ($FinalPD$) assigned to each entity is the maximum of the 12, 24, 36 and 48 month annualised probabilities (Mulder, 2008). That is,

$$FinalPD = Max(AnnPD_{12}, AnnPD_{24}, AnnPD_{36}, AnnPD_{48}) .$$

The maximum of the $AnnPD_i s$ are taken in order to be conservative in assigning ratings.

The rationales for using the above method to derive the final PD are the following:

1. The method assigns the appropriate weight to the variables of the fitted logistic regression model at each point in the PD term structure, which cannot be achieved by having a single iteration only.
2. The method incorporates both short-term and long term risks in the one year PD (it takes into account 1 to 48 month cycles);
3. The PD will be less volatile than just the 12 month cumulative PD, which is more of a point-in time (snapshot) PD compared to the final PD being used.

2.3 Effect size revisited

In this chapter, measures of effect size associated with various statistical techniques for logistic regression modelling have been highlighted. The statistical techniques, measures and significance tests that were discussed earlier in this chapter include Pearson's *r*, for correlation testing, odds ratios for coefficients of regression and their significance, the AUC for the model's ability to discriminate between good and bad credit ratings and overall odds ratio for measuring the fitted model's goodness-of-fit. All of these measures can assist in determining the practical significance of the selected explanatory variables in the logistic regression models used to predict the PDs. This section delves a little bit deeper into the importance of reporting on practical significance by means of effect sizes and how these effect sizes should be interpreted.

2.3.1 Importance of effect sizes

Osteen and Bright (2010) reports that effect size are important because of the reasons that follow.

- In performing power analysis, (where statistical power is the probability of rejecting a false null hypothesis) the statistical power is affected by the estimated effect size (of which the different types are described in more detail in Section 2.3.2 below), the significance level (α), and sample size (n). It basically means that large effect sizes (refer to Table 2.1), high significance levels and bigger samples will improve the statistical power. As a rule of thumb, statistical power ≥ 0.80 is the standard and it indicates a high probability of rejecting a false null hypothesis. Inaccurate estimation of power may lead to wasted resources.
- Knowing the magnitude of an effect allows us to ascertain the practical significance of the statistical test.

Steyn (2009) further reports on the importance of effect sizes in the fields that follow.

- When meta-analysis is conducted, effect size indices are required to combine the results of different studies to an overall measurement, it calls attention to the effect that is associated with the random sampling process in primary studies, it corrects individual study findings for study imperfections, and it examines the variability among previous studies. Hunter and Schmidt (2004) can be consulted for further details on meta-analysis.
- For complete surveys, like for example censuses, where the complete population is studied, effect sizes are essentially the only method to determine the practical importance or significance of results.

For the purpose of this study, focus is placed on magnitude of effect size and reporting these magnitudes practically.

2.3.2 Different types of effect sizes

In terms of the types of statistical analysis that can be performed, we can summarise some of the different types* of effect sizes as follows (for points (1)-(4) below, refer to Osteen and Bright (2010) and for (5) refer to Rice and Harris (2005)):

1. For correlation and regression - Pearson's r, R^2 and Cohen's f^2

1.1 Pearson's correlation coefficient is a statistical measure of the strength of a linear relationship between paired data and by design r is constrained by $-1 \leq r \leq 1$ and furthermore:

- Positive values denote positive linear correlation;
- Negative values denote negative linear correlation;
- A value of 0 denotes no linear correlation;
- The closer the value is to 1 or –1, the stronger the linear correlation.

1.2 The R^2, also known as the coefficient of determination is the proportion of shared variance between 2 or more variables. The value of R^2 is bounded between 0 and 1. R^2 can be interpreted as follows: "R^2(x 100)" percent of the variance in Y can be explained by the variance in the X's

1.3 Cohen's f^2 essentially captures the same relationship as R^2 (where R^2 is the coefficient of determination as defined in 1.2), but in a slightly different form. f^2 is based on R^2 in the sense that it is calculated as:

$$f^2 = \frac{R^2}{1 - R^2}$$.

and f² represents the proportion of the variance explained by one variable to the remaining variance.

2. For logistic regression - Odds Ratios and Pseudo-R^2

2.1 A detailed description of the use of odds ratios in logistic regression models is given in section 2.1.1

2.2 A detailed description of Pseudo-R^2s such as Nagelkerke's R^2, Cox and Snell's R^2 and the Hosmer and Lemeshow's measure is given in Section 2.1.2.1. The Pseudo-R^2 is typically based on the log likelihood for the model with the independent variables ($\ln\hat{L}$ (M_{Full})) compared to the log likelihood for an empty model ($\ln\hat{L}$ ($M_{Intercept}$)) and is given by:

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})}$$.

3. For mean differences - Cohen's *d*, η^2, R^2 and Cohen's f^2

3.1 A measure of effect size that is often used in experiments or studies where one is comparing the mean of one sample to another is that of Cohen's *d* which is discussed in more detail in Section 2.3.3.

3.2 Eta-squared η^2, which is also known as the correlation ratio is often defined as the sums of squares for the effect of interest, divided by the total sums of squares:

$$\eta^2 = SS_{between}/SS_{total.}$$

where the sum of squares between-groups ($SS_{between}$) examines the differences among the group means by calculating the variation of each group's mean around the total mean and where the sum of squares total (SS_{total}) is the sum of the squares of the difference of the dependent variable and its mean. η^2 Is bounded between 0 and 1 and the interpretation thereof is that: "η^2 (x 100)" percent of the variance in Y can be explained by the variance in X.

4. For Crosstabs and Chi-Square - Phi/Cramer's V

4.1 Phi ($\emptyset$) / Cramer's Phi ($\emptyset c$) is a measure of association and is like the effect size equivalent for a Chi Squared statistic. It can also be described as an effect size for categorical data. Cramer's Phi is used when the Chi Squared matrix is bigger than a 2 x 2 matrix. The formula is:

$$\phi_c = \sqrt{\frac{\chi^2}{N(k-1)}}\ .$$

where N = Total number of subjects, k = the smaller of the number of rows or columns and χ^2 is the Chi-squared statistic. Phi/Cramer's V is bounded between 0 and 1 and is interpreted like Pearson's r and R^2

5. For predictive accuracy and discriminative power - ROC area or AUC. Refer to Section 2.1.4 for a detailed description of the ROC area.

*Please note that not all the above-mentioned MES are relevant to this study.

2.3.3 Magnitude of effect

Once the value of the determined effect size is known, (for example the r-value for paired data, or the odds ratio for a variable in the fitted logistic regression model, or the size of the area under the CAP-curve) the question is how to interpret the actual value and its magnitude. The effect sizes are generally broken down into "small", "moderate" or "medium" and "large" values. It is however, important to note that there is not a standard set of rules that can be applied across all the different measures of effect size (MES). What constitutes a small, moderate, or large effect depends on the type of effect size being considered. These three levels are arbitrary and relational; they are guidelines and should not be seen as cut-off values. When quantifying the levels for the types of effect sizes, some of the most often cited references for magnitude of effect include Cohen (1988), Rice and Harris (2005), and Osteen and Bright (2010). Table 2.1 summarises magnitudes of effect for different types of effect sizes.

Table 2.1: Measurable magnitude of effect for different types of effect sizes

Effect Size	Small	Moderate	Large
Pearson's *r*	0.1	0.3	0.5
r^2	0.01	0.09	0.25
η^2	0.01	0.06	0.14
Cohen's R^2	0.01	0.10	0.25
Cohen's *d*	±0.20	±0.50	±0.80
Phi/Cramer's V	0.1	0.3	0.5
Cohen's f^2	0.02	0.15	0.35
AUC	0.55	0.65	0.70
Odds Ratio	1.44	2.47	4.25

Dunst, Hamby and Trivette (2004) report that it is important to select effect size formulas that are applicable to different types of research designs. This yields metrics that can be interpreted in the same manner across the effect sizes. The particular formulas and methods described below are ones generally recommended for computing Cohen's *d* which basically means that all other effect size measures can be expressed in terms of Cohen's *d* through conversion formulas.

Cohen's *d*, the effect-size metric is one of the most widely used measures of magnitude of effect. The formula used for calculating Cohen's *d* is:

$$d = (M_1 - M_2)/\ SD_P$$

where:

M_1 is the mean score of one group of study participants; M_2 is the mean score of a second group of study participants; and SD_P is the pooled standard deviation for both groups of study participants.

There are no agreed upon standards for interpreting the magnitude of effect sizes. However, Cohen's (1977, 1988) original guidelines that $d = 0.20$ is a "small," $d = 0.50$ is a "medium," and $d = 0.80$ is a "large" effect size are still widely cited and used for interpreting magnitudes of effect.

Effect size conversions to Cohen's *d* can be done by using the formulas which follow (for example purposes not all the conversion formulas are included).

- Correlation (Dunst et al. 2004): $d = 2r / \sqrt{1 - r^2}$
- Discriminatory power (Acion et al. 2006, and Rice and Harris 2005):

  $AUC = 1 - 0.5*(1 - d/3.464)^2$ or
  $\mathrm{d} = 3.464(1 - \sqrt{2(1 - \mathrm{AUC})}\ )$

- Odds Ratio (Chinn, 2000): $d = \dfrac{\ln(OR)}{1.81}$

2.3.4 Practical significance

The most important and widely used MES have been defined above and it was shown that the magnitude of the effect can be calculated. One may even know whether or not the inferential test it is based on, is statistically significant. But while conclusions can be drawn, based on statistical significance testing (SST), there may be some shortcomings to be aware of. King (2002) reports that by itself statistical significance testing is inadequate for determining:

- the importance of results; and
- the likelihood of obtaining similar results in the future.

So what exactly does SST assess then? Per definition SST is used to draw inferences about characteristics of populations based on probability samples. The procedures which follow are part of SST.

- Develop a testable hypothesis about a population characteristic ("null hypothesis").
- Draw probability samples from the population.
- Conduct the statistical test, acquiring a probability (p) value
- If the p value is below some criterion (usually arbitrarily set to .05 or .01), then the results are said to be "statistically significant" and the population may not have the characteristic as hypothesised in the null.

In summary it means that SST allows for the calculation of the probability of obtaining sample results based on the assumption that the null hypothesis is true (i.e., assuming that the population is characterised as hypothesised). All of these steps are performed as part of the normal logistic modelling process and have already been discussed in Section 2.1. But what does SST *not* assess? The following may give a good indication of the need to investigate and report on practical significance as well:

- The probability that the null hypothesis is true.
- The probability that the statistical test results were obtained by chance.
- The probability that the same statistical test results will be found in future studies.
- The probability that there is a true effect in the population, which means:
    - SST does not directly measure effect size (i.e., magnitude of observed differences or relationships).
    - Instead, the p value is found by combining the size of the effect with the sample size (the two are confounded). Thus statistical significance may arise due to a large effect, a large sample size, or both.
    - Consequently, results may be "statistically significant" due to a large sample size, but not practically significant due to a small effect (and the converse is true).

It is clear from what has just been highlighted and from what was suggested by Grissom & Kim (2005), Wilkinson and APA (1999), and Steyn (2009) that effect sizes should be used in conjunction with SST (if applicable). That means an effect size measure of the magnitude of differences observed in the data should always be reported, and that this report should include a discussion of the practical significance of the results. Mordock (2000) suggests that, as part of reporting on practical significance, for a value judgment about the effect sizes and its magnitude, the following questions should be asked: is the effect size important, feasible, and practical and does the effect size facilitates decision making?

2.3.5 Reporting guidelines and benefits of reporting effect sizes and its practical significance)

Wilkinson and APA (1999) states that when reporting effect sizes in terms of practical significance, there are three important benefits:

- Meta-analysis: Reporting effect sizes facilitates subsequent meta-analyses incorporating a given report. According to the Merriam-Webster Dictionary (2014), meta-analysis is a quantitative statistical analysis of several separate, but similar experiments or studies in order to test the pooled data for statistical significance.
- Informing subsequent research: Effect size reporting creates a literature in which subsequent researchers can more easily formulate more specific study expectations by integrating the effects reported in related prior studies.
- Interpretation and evaluation of results: Interpreting the effect sizes in a given study facilitates the evaluation of how a study's results fits into existing literature, the explicit assessment of how similar or dissimilar results are across related studies, and potentially informs judgment regarding what study features contributed to similarities or differences in effects.

To summarise everything discussed in Section 2.3, when reporting effect size the following needs to be included:

- Type of effect size;
- Value or magnitude of the effect size;
- Interpretation of the effect size; and
- Practical significance of the effect size.

3 Data analysis and results

3.1 Data analysis

In order to derive a final PD model, (as explained in Chapter 2, Section 2.2) one starts by fitting 48 logistic models to the 1 month before default up to the 48 months before default data sets. In this thesis, for the purpose of illustrating the methods, logistic models will only be fitted to the 1, 24 and 48 months before default data sets.

3.1.1 Data used for modelling

The full modelling data set consisted of 43758 observations. All the exclusions as discussed in Chapter 1 (Section 1.5.2) were done. This data set included some key variables from the financial statements of the companies. These financial variables are used in the credit rating process by credit analysts to assess the companies' credit positions and enable the banks to develop credit rating models based on the historic financial information. The following are definitions of some of these financial variables or ratios in Investopedia (2015).

- Excess Return - Investment returns from a portfolio that exceeds a benchmark or index (for example, the S&P500 index) with a similar level of risk.
- Gearing - The level of a firm's or company's debt related to its equity capital and can also be seen as an indication of a company's extent to which its operations are funded by lenders or investors versus shareholders.
- Market Leverage - The amount of debt used to finance a company's assets. If a company has much more debt than equity, the company is considered to be highly leveraged.
- Turnover - The number of times an asset is replaced or traded during a financial period which may be a month, quarter or year.
- Share - A unit of ownership interest in a company or the company's financial assets.
- Cash-flow Ratio - A measure of how well the current liabilities or financial obligations are covered by the cash flow generated from the counterparty's operations

A snapshot of typical company data between 1995-2008, and a random snapshot of the 1 month before default data set showing different sectors are displayed in Tables 3.1 (a) & (b).

Table 3.1 (a): Extract of the default data set per company (1 Month before default) indicating various financial ratios as observed per month between January 1995 and May 2008.

Company	**Date**	**Sector**	**Sector Name**	**Excess Return**	**Mkt Lev.**	**Gear ing**	**Cashflow Ratio**	**Rel. Size**	**Real TurnOver Growth**	**Debt to Turn Over**	**SharePrice Volatility**	**Share Price High Low**	**Default Indicator**
C473	200009	Industrial	Automotive Parts	0.05	0.27	0.13	0.8969	0.0208	0.075	0.79	0.94	0.67	0
C473	200010	Industrial	Automotive Parts	0.09	0.2	0.13	0.8969	0.0263	0.075	0.79	0.93	0.67	0
C473	200011	Industrial	Automotive Parts	0.22	0.15	0.13	0.8969	0.0263	0.075	0.79	0.9	0.67	0
C473	200012	Industrial	Automotive Parts	0.23	0.14	0.13	0.8969	0.0208	0.075	0.79	0.89	0.67	0
C473	200101	Industrial	Automotive Parts	0.1	0.14	0.13	0.8969	0.0208	0.075	0.79	0.86	0.67	0
C474	199501	Mining & Commodities	Chemicals	0.89	0.58	0.84	0.3750	0.0672	0	0.57	0.49	0.46	0
C474	199502	Mining & Commodities	Chemicals	0.89	0.58	0.84	0.3750	0.0672	0	0.57	0.47	0.46	0
C474	199503	Mining & Commodities	Chemicals	0.77	0.58	0.84	0.3750	0.0672	0	0.57	0.45	0.46	0
C474	199504	Mining & Commodities	Chemicals	0.72	0.58	0.84	0.3750	0.0672	0	0.57	0.46	0.46	0
C474	199505	Mining & Commodities	Chemicals	0.56	0.58	0.84	0.3750	0.0672	0	0.57	0.46	0.46	0
C474	199506	Mining & Commodities	Chemicals	0.57	0.53	0.78	0.5857	0.0735	0.015	0.54	0.42	0.46	0
C474	199507	Mining & Commodities	Chemicals	0.75	0.53	0.78	0.5857	0.0735	0.025	0.54	0.43	0.46	0
C474	199508	Mining & Commodities	Chemicals	0.84	0.53	0.78	0.5857	0.0735	0.03	0.54	0.34	0.46	0
C474	199509	Mining & Commodities	Chemicals	0.89	0.47	0.78	0.5857	0.0735	0.035	0.54	0.2	0.49	0
C474	199510	Mining & Commodities	Chemicals	0.89	0.47	0.78	0.5857	0.0735	0.035	0.54	0.21	0.49	0
C474	199511	Mining & Commodities	Chemicals	0.81	0.46	0.78	0.5857	0.0672	0.035	0.54	0.23	0.51	0
C474	199512	Mining & Commodities	Chemicals	0.7	0.49	0.78	0.5857	0.0672	0.035	0.54	0.29	0.51	0
C474	199601	Mining & Commodities	Chemicals	0.44	0.49	0.78	0.5857	0.0672	0.035	0.54	0.28	0.51	0
C474	199602	Mining & Commodities	Chemicals	0.62	0.45	0.78	0.5857	0.0672	0.035	0.54	0.24	0.52	0
C474	199603	Mining & Commodities	Chemicals	0.57	0.47	0.78	0.5857	0.0611	0.035	0.54	0.29	0.52	0
C474	199604	Mining & Commodities	Chemicals	0.68	0.44	0.78	0.5857	0.0611	0.04	0.54	0.25	0.55	0
C474	199605	Mining & Commodities	Chemicals	0.76	0.42	0.78	0.5857	0.0611	0.04	0.54	0.24	0.59	0
C474	199606	Mining & Commodities	Chemicals	0.79	0.44	0.73	0.5983	0.0798	0.065	0.47	0.3	0.63	0
C474	199607	Mining & Commodities	Chemicals	0.73	0.47	0.73	0.5983	0.0798	0.065	0.47	0.34	0.54	0
C474	199608	Mining & Commodities	Chemicals	0.73	0.47	0.73	0.5983	0.0798	0.065	0.47	0.25	0.49	0
C474	199609	Mining & Commodities	Chemicals	0.58	0.45	0.73	0.5983	0.0735	0.06	0.47	0.19	0.49	0
C474	199610	Mining & Commodities	Chemicals	0.6	0.44	0.73	0.5983	0.0735	0.06	0.47	0.22	0.45	0
C474	199611	Mining & Commodities	Chemicals	0.49	0.49	0.73	0.5983	0.0735	0.06	0.47	0.37	0.51	0

Table 3.1 (b): Random Extract of the default data set, showing different sectors (1 Month before default) and indicating various financial ratios as observed per month between January 1995 and May 2008.

Company	Date	Sector	Sector Name	Excess Return	Mkt Lev	Gearing	Cash flow Ratio	Rel Size	Real TO Growth	Debt to TO	Share Price Volatility	Share Price High Low	Default Indicator
C14	199502	Industrial	Construction & Materials	0.76	0.19	0.11	0.7290	0.5831	0	0.78	0.34	0.61	0
C32	199801	TMT	Media	0.34	0.05	0.15	0.8115	0.0432	0.145	0.56	0.63	0.04	0
C49	199511	Consumer Products	Beverages	0.44	0.17	0.21	0.8115	0.6438	0.03	0.12	0.2	0	0
C54	199810	Consumer Products	Food Producers	0.17	0.18	0.25	0.0764	0.1871	0.13	0.91	0.89	0.47	0
C74	199502	Mining & Commodities	Industrial Metals	0.85	0.45	0.33	0.4648	0.9263	0.01	0.26	0.95	0.75	0
C85	200007	TMT	Household Goods	0.16	0.74	0.73	0.0028	0.2278	0.085	0.91	0.9	0.68	0
C169	200403	Services	Support Services	0.37	0.57	0.93	0.0080	0	0.18	0.93	0.58	0.98	0
C222	199712	Services	Pharmaceutics & Biotechnology	0.64	0.66	0.25	0.9114	0.5135	0.05	0.74	0.07	0.03	0
C228	200110	Services	General Retailers	0.45	0.86	0.95	0.2722	0.2031	0.155	0.56	0.84	0.9	0
C235	200110	TMT	Software & Computer Services	0.61	0.08	0.16	0.7155	0.0102	0.11	0.31	0.68	0.9	0
C239	200803	TMT	Electrical Equipment	0.24	0.02	0.17	0.9702	0.0102	0	0.17	0.59	0.3	0
C256	199603	Mining & Commodities	Chemicals	0.32	0.64	0.68	0.2530	0.8411	0.045	0.62	0.31	0.46	0
C260	200602	Mining & Commodities	Mining	0.11	0.11	0.16	0.0080	0.0208	0.34	0	0.77	0.8	0
C268	200410	Services	Health Care Equipment	0.36	0.48	0.6	0.6889	0.6942	0.035	0.54	0.04	0.07	0
C289	199904	Industrial	Automobiles & Parts	0.16	0.94	0.96	0.0270	0.0263	0	0.67	0.88	0.77	0
C311	200212	Industrial	General Industrials	0.94	0.61	0.46	0.3019	0.2363	0.06	0.85	0.52	0.49	1
C356	200006	Industrial	Automobiles & Parts	0.31	0.63	0.44	0.1983	0.3263	0.105	0.85	0.57	0.27	0
C390	200706	TMT	Media	0.49	0.14	0.22	0.7837	0.5712	0	0.61	0.07	0.18	0
C402	200106	Services	General Financial	0.01	0.81	0.26	0.7290	0.1415	0.155	0.96	0.97	0.98	0
C405	200305	Services	Travel & Leisure	0.87	0.21	0.18	0.8969	0.1638	0.035	0.12	0.41	0.23	0
C408	200704	Services	Industrial Transportation	0.47	0.32	0.12	0.5240	0.1715	0.18	0.26	0.19	0.49	0
C438	199909	TMT	Technology Hardware	0	0.76	0.04	0.1032	0.0432	0.03	0.57	0.98	0.63	0
C471	199812	Services	Other	0.33	0.37	0.52	0.3968	0.4263	0.02	0.2	0.77	0.42	0
C501	200206	Mining & Commodities	Oil & Gas Producers	0.77	0.39	0.59	0.7425	0.9702	0.115	0.8	0.59	0.46	0
C507	200712	Services	Support Services	0.4	0.03	0.03	0.6626	0.0318	0.13	0.82	0.36	0.04	0
C530	199511	Consumer Products	Beverages	0.5	0.11	0.12	0.8115	0.5022	0.015	0.13	0.13	0.97	0
C549	199807	Industrial	Industrial Engineering	0.99	0.55	0.88	0.0010	0.2622	0.17	0.65	0.99	0.99	0
C566	200804	TMT	Software & Computer Services	0.46	0.21	0.97	0.1983	0.3263	0.165	0.76	0.46	0.76	1
C571	199503	Mining & Commodities	Forestry & Paper	0.81	0.51	0.35	0.3430	0.0863	0.06	0.45	0.45	0.3	0

Table 3.2 below summarises the descriptions of the data-fields / variables and variable abbreviations in the modelling data sets.

Table 3.2: Description of the data-fields / variables and variable abbreviations.

Variables	Description
Datestamp	1995/01 - 2008/04
Comp	Company Code: C1 - C572 (Number of Companies)
Main_Sector	Customer Products
	Services
	Industrial
	TMT
	Mining and Commodities
Sector_Name	Automobiles & Parts
	Beverages
	Blank
	Chemicals
	Construction & Materials
	Electricity
	Electronic & Electrical Equipment
	Equity Investments Instruments
	Fixed Line Telecommunications
	Food & Drug Retailers
	Food Producers
	Forestry & Paper
	General Financial
	General Industrials
	General Retailers
	Health Care Equipment & Services
	Household Goods
	Industrial Engineering
	Industrial Metals
	Industrial Transportation
	Leisure Goods
	Media
	Mining
	Mobile Telecommunications
	Oil & Gas Producers
	Other
	Personal Goods
	Pharmaceuticals & Biotechnology
	Software & Computer Services
	Support Services
	Technology Hardware & Equipment
	Travel & Leisure
ExcessReturn	Excess Return on Share Price
Gearing	Using borrowed funds or debt to supplement existing funds for investment
CFtoTD	Cashflow Ratios
mktLeverage	Market Leverage
DtoTOver	Debtors / Turnover
realTOverGrowth	Real Turnover Growth
relSize	Relative Size Indicator (Listed Companies)
Share_High_Low	Share Price High vs. Low
stDevSharePrice	Standard deviation of Share Price (Share Price Volatility)
Indicator	Flagged a default = 1 and non-default = 0 (Total of 90 defaults occurred)

Table 3.3 displays the descriptive statistics of the complete data set. The table shows that there were no missing values. That is, the data set was complete.

Table 3.3: Descriptive statistics of the complete data set

	Mean	Std Error	Std Dev	Variance	Range	Min	Max	Sum	Count (N)
ExcessReturn	0.4954	0.0014	0.2883	0.0831	1	0	1	21678	43758
mktLeverage	0.4944	0.0014	0.2885	0.0832	1	0	1	21633	43758
Gearing	0.4950	0.0014	0.2882	0.0831	1	0	1	21659	43758
CFtoTD	0.3182	0.0014	0.2891	0.0836	1	0	1	13922	43758
relSize	0.3511	0.0014	0.2826	0.0799	0.9851	0	0.9851	15366	43758
realTOverGrowth	0.0834	0.0004	0.0816	0.0067	0.3400	0	0.3400	3648	43758
DtoTOver	0.4962	0.0014	0.2883	0.0831	1	0	1	21711	43758
stDevSharePrice	0.4935	0.0014	0.2891	0.0836	1	0	1	21594	43758
Share_High_Low	0.4939	0.0014	0.2885	0.0832	1	0	1	21614	43758
ExcessReturn_S	0.4960	0.0014	0.2887	0.0833	1	0	1	21706	43758
stDevSharePrice_S	0.4935	0.0014	0.2894	0.0838	1	0	1	21594	43758
Indicator	0.0021	0.0002	0.0453	0.0021	1	0	1	90	43758

Key: Refer to Table 3.2 for explanations on variable abbreviations

The descriptive statistics in Table 3.3 assisted in identifying outliers and missing values. One could observe there were no missing values and that the data set was complete. The range of the predictor or independent variables (which are all financial ratios) was between 0 and 1, except for Real Turn-over Growth (realTOverGrowth) where the range was even smaller. In order to look at the dispersion of the different variables, in other words how widely spread the observations in the specific variable is from the mean, the standard deviation is a good measure that can be used. The standard deviation in Table 3.3 indicates that most of the predictor variables show variation or movement over the observations. For example, for Share Price High vs. Low (Share_High_Low) this may be interpreted as the presence of a range of differences in the share prices between the various companies over the time-period of investigation. If the spread of the data is close to the mean, the standard deviation will be small, giving an indication of few or no outliers or influential observations. For the purpose of this thesis, not much focus had to be placed on meeting the assumptions of normality as logistic regression models were used for variable selection and predicting the probability of default. Therefore the skewness and kurtosis statistics for the explanatory variables were not interpreted.

A summary of the data sets with snapshots at 1, 12, 24, 36 and 48 months before default were as follows:

- Data set 1 consisted of all observations in the database up to one month before default was observed (572 companies, 43 758 observations & 90 Defaults);
- Data set 12 consisted of all observations in the database up to 12 months before default was observed (517 Companies, 37 743 Observations & 79 Defaults);
- Data set 24 consisted of all observations in the database up to 24 months before default was observed (474 Companies, 31 783 Observations & 68 Defaults);
- Data set 36 consisted of all observations in the database up to 36 months before default was observed (424 Companies, 26 360 Observations & 58 Defaults); and
- Data set 48 consisted of all observations in the database up to 48 months before default was observed (378 Companies, 21 520 Observations & 44 Defaults).

The 0 or 1 values of the default indicator variable assigned to a company each month were based on the assessment of the company's performance in terms of the Basel II definition of default and the Bank's specific portfolio strategies as was described in Chapter 1 (Section 1.5.3).

3.1.2 Correlation of variables and cluster analysis

Table 3.4 displays the Pearson's correlation matrix of the explanatory variables. The matrix was constructed from the complete data set (January 1995 to May 2008) using Excel 2007. From a business experience point of view it should be declared which coefficient thresholds or values are considered as limits for weak, moderate and high correlations. It can for example, be declared that two variables with a $|r| > 0.4$ are highly correlated and need to be further investigated for this by means of the VIF as was described in Chapter 2 (Section 2.1.3.1), in order to see if there's any collinearity issues before considering both variables for inclusion into a model.

Table 3.4: The Pearson's correlation matrix of the explanatory variables in the complete data set. (1 month before default)

	Excess Return	mkt Leverage	Gearing	CF to TD	relSize 1	Real TOver Growth	DtoT Over	stDev Share Price	Share_ High _ Low
Excess Return	1.0000	-0.2511	-0.0271	0.1473	0.1324	-0.0689	-0.0492	-0.2938	-0.0687
mktLeve rage	-0.2511	1.0000	0.4698	-0.5212	-0.2953	0.0542	0.0714	0.2360	0.1519
Gearing	-0.0271	0.4698	1.0000	-0.3628	-0.0178	0.0503	0.0405	0.0873	0.1039
CFtoTD	0.1473	-0.5212	-0.3628	1.0000	0.2546	-0.2255	-0.1623	-0.2661	-0.2145
relSize	0.1324	-0.2953	-0.0178	0.2546	1.0000	-0.2403	-0.1546	-0.3581	-0.3438
realT Over Growth	-0.0689	0.0542	0.0503	-0.2255	-0.2403	1.0000	0.0600	0.1889	0.2116
DtoT Over	-0.0492	0.0714	0.0405	-0.1623	-0.1546	0.0600	1.0000	0.1013	0.1001
stDev Share Price	-0.2938	0.2360	0.0873	-0.2661	-0.3581	0.1889	0.1013	1.0000	0.5231
Share_ High_ Low	-0.0687	0.1519	0.1039	-0.2145	-0.3438	0.2116	0.1001	0.5231	1.0000

Key: Refer to Table 3.2 for explanations of variable abbreviations

The guidance of Cohen (1988) for magnitude of effect sizes and Cohen's $d = 2r/\sqrt{1-r^2}$ was used (refer to Chapter 2, (Table 2.1)) to obtain the values and size indicators in Table 3.5, and to declare correlations in Table 3.4 as weak, moderate and large. The moderate and large correlations are shaded. For example, according to the guidance of Cohen the variables Market Leverage (mktLeverage) and Share Price High vs. Low (Share_High_Low) has a weak correlation while the variables Market Leverage (mktLeverage) and Cashflow Ratios (CFtoTD) are highly correlated (see Table 3.5). The highly correlated variables could give rise to the problem of multi-collinearity if both variables were included in the model. In the analysis only one of a pair of moderately or highly correlated explanatory variables was included in the model to avoid the problem of multi-collinearity

Table 3.5: The conversion of Pearson's correlation coefficients to Cohen's *d* values and the corresponding indication of the magnitude of the effect size

Pearson's *r*	r^2	Cohen's *d*	Effect	Variance in one variable accounted for by variance in other variable
0.0405	0.0016	0.0811	Small/Weak	0%
0.1519	0.0231	0.3074	Small/Weak	2%
-0.2255	0.0509	-0.4629	Moderate	5%
0.2744	0.0753	0.5707	Moderate	7%
-0.3438	0.1182	-0.7322	Large	12%
-0.3628	0.1316	-0.7787	Large	13%
0.4698	0.2207	1.0644	Large	22%
-0.5212	0.2716	-1.2214	Large	27%
0.5231	0.2736	1.2275	Large	27%

As part of correlation analysis, the explanatory variables were also screened by using cluster analysis of the variables based on the correlation among the variables (refer to Section 2.1.3.1 for a description of cluster analysis). The results for the one month before default data sets are displayed in Table 3.6. From this table it can be seen that for example, Market Leverage, Gearing and Cash Flow are clustered together (Cluster 2) which indicates correlation among these three variables and this is basically a confirmation of what was already observed in the Pearson's correlation matrix in Table 3.4.

Table 3.6: The groups of correlated explanatory variables - cluster analysis based on the correlations among the variables.

1 Month before default data set				
4 Clusters		**R-squared with**		**1-R**2**
Cluster	**Variable**	**Own Cluster**	**Next Closest**	**Ratio**
Cluster 1	**RelSize**	48%	6%	55%
	RealT Over Growth	25%	2%	76%
	StDev Share Price	61%	9%	42%
	Share_ High_ Low	61%	4%	40%
Cluster 2	**MktLeve rage**	71%	7%	31%
	Gearing	57%	1%	43%
	CFtoTD	62%	12%	43%
Cluster 3	**DtoT Over**	100%	2%	0%
Cluster 4	**Excess Return**	100%	4%	0%

Key: Refer to Table 3.2 for explanations on variable abbreviations

If cluster analysis is used to assist in the variable selection process, the variable with a low $1-R^2$ ratio should be the one to select. In Cluster 1, the variable likely to exclude based on its "higher" $1-R^2$ ratio of 76%. This is however not a "cast in stone" rule and two variables from the same cluster can be included into the model if they have a low $1-R^2$ ratio and it makes business sense to consider both.

Although the assessments of correlation and magnitude of effect size was displayed for the 1 month before default data set only, the processes above were performed on all the data sets.

3.1.3 Modelling data sets (month 1 – 48 before default)

The following nine independent variables from the list in Table 3.2 were considered for inclusion in the PD models:

- m1 = Excess Return on Share Price;
- m2 = Market Leverage;
- m3 = Gearing (Net Fair Value Debt / Net Tangible Assets);
- m4 = Cash Flow Ratio (Cash Flow / Net Fair Value Debt);
- m5 = Size Indicator (Minimum of Market Cap and Adjusted Turn Over);
- m6 = Real Turnover Growth indicator;
- m7 = Debtors / Turnover;
- m8 = Share Price Volatility; and
- m9 = (Three year high – Three Year Low)/ Share Price Three year low.

The nine variables were renamed m1 - m9 because the original variable names were long and caused statistical analysis summary tables to be too large. These nine variables were the initial set of explanatory variables for fitting logistic models to the 1 - 48 months before default data sets. The data analysis was performed using IBM SPSS Statistics version 20, STATISTICA version 11, StatSoft, Inc. (2012), SAS Enterprise Guide version 5.1 and MS Excel 2010. The output is in the sections that follow.

3.2 Fitting the full logistic regression model to the data

In order to apply the stepwise variable selection procedure one need to check model assumptions, outliers etc. first while all nine variables are still in the model. Once the required assumptions are met, one can perform the stepwise selection procedure.

3.2.1 Descriptive statistics and correlation analysis

The variable selection process will be demonstrated by fitting the logistic model to the 48 month before default data set and is described here in detail, in order to illustrate the similar procedures followed for all other data sets. Descriptive statistics were generated for the 9 variables.

Table 3.7: Descriptive statistics for the nine variables to be included in the full model fitted to the 48 month before default data set.

Variable	Mean	Std Dev	Std Error	Variance	Min	Max	Range	N	N Miss
m1	0.5179	0.2832	0.0019	0.0802	0	0.9900	0.9900	21520	0
m2	0.5056	0.2819	0.0019	0.0795	0	0.9900	0.9900	21520	0
m3	0.4926	0.2779	0.0019	0.0772	0	0.9900	0.9900	21520	0
m4	0.3185	0.2816	0.0019	0.0793	0	0.9850	0.9850	21520	0
m5	0.3781	0.2875	0.0020	0.0827	0	0.9851	0.9851	21520	0
m6	0.0795	0.0773	0.0005	0.0060	0	0.3400	0.3400	21520	0
m7	0.4945	0.2811	0.0019	0.0790	0	1.0000	1.0000	21520	0
m8	0.4830	0.2886	0.0020	0.0833	0	0.9900	0.9900	21520	0
m9	0.4868	0.2820	0.0019	0.0795	0	1.0000	1.0000	21520	0

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low

The descriptive statistics in Table 3.7 assisted in identifying outliers and missing values. One could observe there were no missing values and that the data set was complete. The range of the predictor or independent variables (which are all financial ratios) was between 0 and 1, except for Real Turn-over Growth (m6) where the range was even smaller. Again not much focus had to be placed on meeting the assumptions of normality as logistic regression models were used for variable selection and predicting the probability of default. From Table 3.7 it can be seen that the predictor variables were all within range, but one cannot really observe whether there are outliers present. Therefore, once the model is fitted to the data set, Pearson and Deviance Residual plots, Leverage plots and DFBeta plots can be analysed as they give an indication of outliers or extreme observations and whether these outliers are influential (see Section 3.1.4.2).

Table 3.8: The Pearson's correlation Matrix of the explanatory variables in the 48 Month before default data set.

	m1	m2	m3	m4	m5	m6	m7	m8	m9
m1	1	-0.1949	0.0112	0.1057	0.1003	-0.0421	-0.0223	-0.3163	-0.0790
m2	-0.1949	1	0.5091	-0.5243	-0.2968	-0.0023	0.0565	0.2174	0.1286
m3	0.0112	0.5091	1	-0.3980	-0.0328	0.0798	0.0444	0.0861	0.1019
m4	0.1056	-0.5243	-0.3980	1	0.2047	-0.1680	-0.1476	-0.2392	-0.2022
m5	0.1003	-0.2968	-0.0328	0.2048	1	-0.1941	-0.1380	-0.3300	-0.3010
m6	-0.0421	-0.0023	0.0798	-0.1680	-0.1941	1	0.0416	0.1889	0.2170
m7	-0.0223	0.0565	0.0444	-0.1476	-0.1380	0.0416	1	0.0544	0.0726
m8	-0.3163	0.2174	0.0861	-0.2392	-0.3300	0.1889	0.0544	1	0.4843
m9	-0.0790	0.1286	0.1019	-0.2022	-0.3010	0.2170	0.0726	0.4843	1

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low. Moderate to high correlations (see Table 3.9) are shaded.

The correlation matrix of the 9 variables in the 48 month before default data set (Table 3.8) and conversion of Pearson's correlation coefficients to Cohen's d values (Table 3.9) indicated that the variables Market Leverage (m2) and Cash Flow Ratio (m4), as well as Market Leverage (m2) and Gearing (m3) were highly correlated (where Pearson's r = -0.521 and Cohen's d = -1.2305 for m2 and m4, and Pearson's r = 0.5091 and Cohen's d = 1.1827 for m2 and m3,) when using a threshold of |r|>0.4. When using this criterion, it could be considered to remove m2, m3 or m4 or even two of them from the model.

Table 3.9: The conversion of Pearson's correlation coefficients to Cohen's d values and the corresponding indication of the magnitude of the effect size.

Combinations	**Pearson's *r***	**r²**	**Cohen's *d***	**Effect**	**Variance in one variable accounted for by variance in other variable**
m2 & m3	.5091	0.2591	1.1827	Large	26%
m2 & m4	-.5243	0.2746	-1.2305	Large	27%
m8 & m9	.4843	0.2343	1.1062	Large	23%
m3 & m4	-.3980	0.1584	-0.8677	Large	16%
m5 & m8	-.3300	0.1089	-0.6992	Moderate	11%
m1 & m8	-.3163	0.0999	-0.6661	Moderate	10%

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low.

Cluster analysis was also performed (discussed in Section 3.1.2) as a further indication of which variables are correlated, and which variable from the different clusters to potentially include in the model to be fitted. The results are in Table 3.10. In Cluster 2, the variables Market Leverage (m2), Gearing (m3) and Cash Flow Ratio (m4) are correlated and typically only one of them should be included into the model. From Cluster 1, one would maybe want to include variables Share Price Volatility (m8) or Share Price High/Low (m9) in the model.

Table 3.10: The groups of correlated explanatory variables - cluster analysis based on the correlations among the 9 variables.

4 Clusters			**R-squared with**	**1-R**2**
Cluster	**Variable**	**Own Cluster**	**Next Closest**	**Ratio**
Cluster 1	**m5**	44%	5%	59%
	m6	25%	1%	76%
	m8	60%	10%	44%
	m9	60%	3%	42%
Cluster 2	**m2**	72%	6%	30%
	m3	61%	1%	39%
	m4	63%	9%	41%
Cluster 3	**m7**	100%	1%	0%
Cluster 4	**m1**	100%	4%	0%

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low.

3.2.2 Checking model assumptions

The following SAS output (Table 3.11) was obtained when the Logistic Procedure in SAS Enterprise Guide Version 5.1 was used to fit the logistic regression model with all 9 the variables to the 48 month before default data set.

Table 3.11: Model fit statistics of the logistic regression model with all 9 variables fitted to the 48 month before default data.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates (9)**
AIC	634.854	606.357
SC	642.831	686.124
-2 Log L	632.854	586.357

No. of observations = 21520

The comparison of the model fit statistics of the model with an intercept only with those of the model with an intercept and all 9 variables, the model with the smallest Akaike Information Criterion (AIC) fits better. Therefore the model with intercept and covariates was considered the better fit. There is however an inherent warning in the model fit statistics. Swanson *et al.* (2010) report that in most cases preference would be to a model that has the fewest parameters to estimate provided that the candidate models are correctly specified. This is called the most "parsimonious" model of the set. The Schwarz Criterion (SC) would pick a more parsimonious model than what the AIC might suggest. What is therefore clear from Table 3.11 is that although the AIC indicates that the model with intercept and 9 covariates was a better fit than the intercept only model, SC suggests that a model with fewer covariates would be an even better fit.

a) Linearity

The assumption when fitting the logistic regression model to the data was that the logit(p) is linearly related to the predictor variables. It is therefore crucial to check this assumption. As discussed in Chapter 2 (Section 2.1.2.2), the Box-Tidwell transformation and test is used to check this assumption of linearity. For each continuous predictor variable (X), Xln(X) was calculated and the logistic model with 9 Xln(X) predictor variables fitted to the 48 months before default data. The results are included as can be seen in Table 3.12. If a Xln(X) is significant, then the assumption of linearity is violated. In other words, if the coefficient of variable Xln(X) is statistically significant ($p<0.05$), there is evidence of non-linearity in the relationship between Logit(p) and that specific X variable. Table 3.12 shows logit(p) is not linearly related with Real Turnover Growth Indicator (m6) as p=0.0229<0.05 and Debtors/Turnover (m7) as p=0.0001<0.05.

Table 3.12: The output from fitting the logistic model with 9 Xln(X) variables to the 48 month before default data.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	5.577	1.0831	26.5125	<.0001
m1×ln(m1)	1	-0.623	1.477	0.178	0.6731
m2×ln(m2)	1	0.2755	1.7798	0.024	0.877
m3×ln(m3)	1	1.2755	1.7398	0.5374	0.4635
m4×ln(m4)	1	-2.0758	1.4659	2.0053	0.1567
m5×ln(m5)	1	2.8197	1.9819	2.0242	0.1548
m6×ln(m6)	1	4.8581	2.1356	5.1747	0.0229
m7×ln(m7)	1	-6.3435	1.5456	16.8437	<.0001
m8×ln(m8)	1	-1.4303	1.5586	0.8421	0.3588
m9×ln(m9)	1	-1.8443	1.6789	1.2068	0.272

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low

According to Hair *et al*. (1998), the most direct approach when non-linear relationships are detected is to transform the predictor variables to achieve linearity or include the Xln(X) of the variables in the model. It may however not be necessary to perform any transformations on these variables if the correlation analysis and stepwise selection process give an indication that the variables should be removed from the fitted model.

b) Independence of the responses/observations

Another important assumption that needs to be met by the data when fitting logistic regression models is uncorrelated responses. The violation of this assumption will potentially be a problem if the best predictor of the next observation is based upon some function of the current observation (or a prior observation). This may typically happen if there are “cycles” or periods where the probability of default increased significantly and probability of default for a specific month is highly correlated with the previous month’s probability of default. If this was the case, the impact of these “cycles” needs to be investigated or taken into account in selecting variables for the models. Therefore auto-correlation needed to be tested, to assess the independence of observations (especially where these observations were for the same company for consecutive months). The Durbin-Watson (D-W) test was performed on the residuals from the models. As reported by Hair *et al*. (1998), the test criteria is that if a value of the D-W statistic = 2, it indicates no autocorrelation. The output also includes a p-value. Conventionally, if $p < 0.05$, then the residuals are significantly correlated whereas $p > 0.05$ provides no evidence of correlation.

Table 3.13: The Durbin-Watson test statistic for checking the presence of autocorrelation in the residuals of the logistic regression model with all 9 variables fitted to the 48 month before default data.

Model	Durbin-Watson	Sig.
48 Months	2.004[a]	0.0875

a. Predictors: (Constant), m1, m2, m3, m4, m5, m6, m7, m8, m9
Sig.=p-value
Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low.

Upon fitting the logistic model with all 9 variables to the 48 month before default data, a D-W = 2.004 and a p-value = 0.08753 > 0.05 were obtained, which means that there was no indication of autocorrelation and the assumption of independence of responses was not violated.

c) Uncorrelated explanatory variables (multi-collinearity)

As dependencies among the explanatory variables cause parameter estimates to be unstable, care should be taken not to include highly correlated predictor or explanatory variables into the fitted model. To perform multi-collinearity diagnostics when variables that have already displayed evidence of strong correlation (see Tables 3.8 to 3.10) are still in the model would make less sense than to remove some of these variables first and then assess if there are still any multi-collinearity problems with the variables left in the fitted model. The checking of multi-collinearity will therefore be revisited once the variable selection process (Pearson's correlation, cluster analysis and stepwise selection) has been completed and the model refitted with the remaining variables (Refer to Section 3.3.2).

3.2.3 Check for outliers and influential observations

In order to check for outliers and influential observations, the residuals from fitting the logistic model were examined. As reported by Hair *et al.* (1998), residuals are useful in identifying observations that are not explained well by the model. Hair *et al.* (1998) also reports that the Pearson residuals are components of the Pearson chi-square statistic and deviance residuals are components of the deviance goodness-of-fit statistics in Table 3.11. The logistic model with all 9 variables was fitted to the 48 month before default data and the Pearson residuals and Deviance residuals obtained from fitting the model was used to graphically check for outliers. This basically means that potential outliers were identified through a visual inspection, where cases that were far away from most of the others were highlighted. The results are displayed in Figure 3.1. The index plots of the Pearson residuals and the Deviance residuals in Figure 3.1 indicated that 1 case between no's 10000 and 15000 is far away from most of the other observations. This is a point that may need particular attention as it has a very high Pearson residual. This case appears to be an outlier and could therefore be poorly accounted for by the model. Figure 3.1 also includes a graph of Leverage vs. Case (Observation) number for checking for influential observations. Again a visual inspection of this plot was performed in order to identify potential influential observations. The leverage of a specific observation is an overall measure of the potential influence that this observation can have on the analysis and these are usually referred to as the "hat-values" which is calculated from the diagonal elements of the design matrix $\mathbf{X_d}$ in Chapter 2 (Section 2.1.1). The index plot of leverage suggests that a case close to no. 10000 is an influential observation in the data.

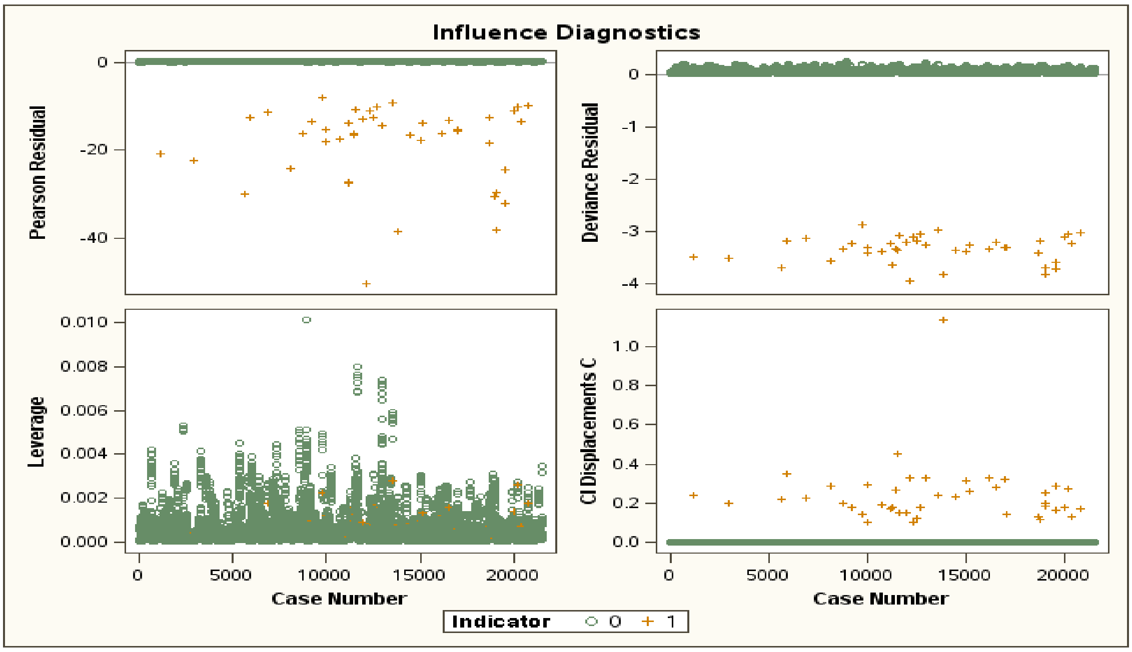
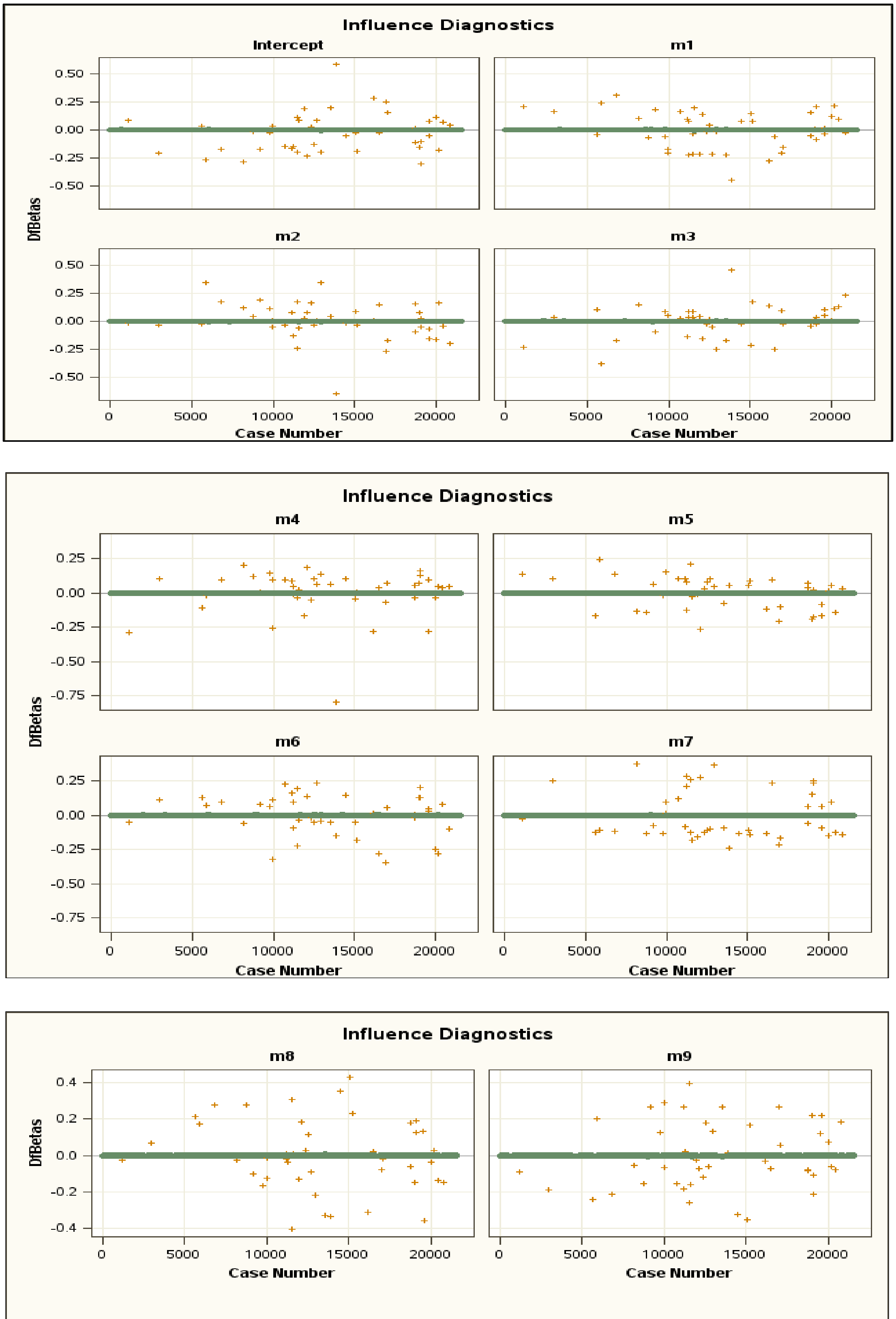


Figure 3.1: Residual and leverage plots from fitting the logistic model with 9 variables to the 48 months before default data.

Hair *et al.* (1998) reports that the DFBETA is a measure of the change in a regression coefficient when an observation is omitted from the regression analysis. The index plots of DFBETAS from fitting the logistic model with 9 variables to the 48 months before default data are displayed in Figure 3.2. The plots indicate that a case close to no. 13000 may be causing instability in parameter estimates for Excess Return of Share Price (m1), Market Leverage (m2), Gearing (m3) and Cash Flow Ratio (m4). This may also be an influential observation,



Key: m1 = Excess Return of Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low

Figure 3.2: DFBETA plots from fitting the logistic model with 9 variables to the 48 months before default data.

Once an outlier has been detected it may be worth removing that observation from the data, and refitting the model to the data in order to see whether this observation has any influence on the goodness-of-fit of this model. According to Hair *et al*. (1998) one would not really expect to identify many poorly fit or influential points when the model seems to fit well on overall goodness-of-fit tests like the Hosmer and Lemeshow test. If this test (refer to Section 3.3.2.2) shows that the model does not fit well it may be worth to revisit the few observation points identified as potential outliers or influential point. It is maybe also important to note the fact that this was a very big data set (21520 observations). This means that the small number of outliers (one or two) wouldn't have much influence (if any) on the goodness of fitting the logistic regression model and therefore the outliers were not removed, but these observations were just left in the data set for modelling.

3.2.4 Stepwise Variable selection

Once the full model was fitted, and it was clear that the full model fitted better than the empty model, it was also observed that some of the variables are highly correlated and some of these variables may have to be removed from the model. In order to decide on the optimal set of variables, various variable selection techniques and diagnostic tests as discussed in Chapter 2 (Section 2.1.3) were employed and these quantitative techniques were then combined with business experience to come up with the optimal set of variables for the model to be fitted to the 48 month before default data set. This process was repeated for all other data sets (1-47 months before default data sets) as well, however, the results are not shown here. The 48 months before default data set was used as an illustrative example of the process of logistic model fitting.

The stepwise variable selection procedure described in Chapter 2 (Section 2.1.3.2) was used next to select the optimal set of explanatory variables. The significance level at which variables entered or exited the model was 5%. The result of the application procedure is displayed in Table 3.14. These are from fitting the logistic model to the 48 month before default data set.

Table 3.14: The results of fitting the logistic model with all 9 variables to the 48 month before default data set, using the stepwise selection procedure (entry and removal significance level was at 5%).

Variables not in the model			Chi-square	df	Sig.
Step 2[a]	Variables	m3	.077	1	.781
	Overall Statistics		.077	1	.781
Step 3[b]	Variables	m3	.064	1	.801
		m9	.137	1	.711
	Overall Statistics		.215	2	.898
Step 4[c]	Variables	m1	.241	1	.623
		m3	.120	1	.729
		m9	.087	1	.767
	Overall Statistics		.453	3	.929
Step 5[d]	Variables	m1	.234	1	.629
		m3	.158	1	.691
		m6	.573	1	.449
		m9	.049	1	.825
	Overall Statistics		1.029	4	.905
Step 6[e]	Variables	m7	1.834	1	.176
		m1	.003	1	.954
		m3	.167	1	.682
		m6	.721	1	.396
		m9	.156	1	.693
	Overall Statistics		2.855	5	.722

Sig.=p-value; Overall Statistic – Chi-square test statistics for the significance of variable(s) ;
Key: m1 = Excess Return of Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low

Table 3.14 shows that for the 48 month before default data set, the explanatory variables that were significant or left in the model were:

- m2 = Market Leverage;
- m4 = Cash Flow Ratio (Cash Flow / Net Fair Value Debt);
- m5 = Size Indicator (Minimum of Market Cap and Adjusted Turn Over); and
- m8 = Share Price Volatility.

It was however decided that based on the Pearson’s correlation coefficients, Cohen’s d values and the Cluster analysis to remove Market Leverage (m2) from the model as well.

3.3 Fitting the logistic regression model for selected variables to the 48 month before default data set

3.3.1 Checking correlation and model assumptions

The maximum absolute correlation |r| among the 3 variables (Cash Flow Ratio (m4), Size Indicator (m5) and Share Price Volatility (m8)) remaining in the logistic model fitted to the 48 month before default data set was 0.3300 as can be seen in Table 3.15, which is a moderate effect size according to Cohen’s d = -0.6992 as can be seen from Table 3.16.

Table 3.15: The Pearson's correlation matrix of the remaining 3 variables from fitting the logistic model to the 48 months before default data.

	m4	m5	m8
m4	1	0.2047	-0.2392
m5	0.2047	1	-0.3300
m8	-0.2392	-0.3300	1

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility.

Table 3.16: The conversion of Pearson's correlation coefficients to Cohen's d values and the corresponding indication of the magnitude of the effect size.

Combinations	Pearson's *r*	r^2	Cohen's *d*	Effect	Variance in one variable accounted for by variance in other variable
m5 & m8	-.3300	0.1089	-0.6992	Moderate	11%

Key: m5 = Size Indicator and m8 = Share Price Volatility

The assumption of linearity between the logit(p) and the three individual variables was met by the data, as the smallest p-value was 0.0797>0.05 for the significance of Xln(X) of the Cash Flow Ratio (m4) in Table 3.17.

Table 3.17: The output from fitting the logistic model with 3 Xln(X) variables to the 48 month before default data.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	0.0037	0.0013	2.9510	0.0032
m4×ln(m4)	1	0.0067	0.0029	2.3330	0.0797
m5×ln(m5)	1	-0.0043	0.0032	-1.3407	0.1800
m8×ln(m8)	1	0.0046	0.0030	1.5141	0.1300

From the test performed and the results in Table 3.18, there was no indication of autocorrelation (D-W=2 and p-value=0.0864) and the assumption of independence of responses was not violated.

Table 3.18: The Durbin-Watson test statistic for checking the presence of autocorrelation in the residuals of the remaining 3 variables from fitting the logistic model to the 48 months before default data.

Model	Durbin-Watson	Sig.
48	2.004[a]	0.0864

a. Predictors: (Constant), m8, m4, m5;

Sig.=p-value

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility.

The potential effect of correlated independent variables in the model was further investigated using multi-collinearity diagnostics (see Section 2.1.3).

Table 3.19 displays the variance inflation factors (VIF) and the tolerance levels of the 3 variables: Cash Flow Ratio (m4), Size Indicator (m5) and Share Price Volatility (m8), when fitting the logistic models to the 48 months before default data set. This table shows that there were no multi- collinearity issues in the fitted model as none of the variables had tolerance levels less than 0.1 or VIF values greater than 10.

Table 3.19: Variance inflation factors (VIF) and tolerance levels of the remaining 3 variables from fitting the logistic model to the 48 months before default data.

Model	Collinearity Statistics	
	Tolerance	VIF
m4	.925	1.081
m5	.874	1.144
m8	.860	1.162

Dependent Variable: Default Indicator

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility.

The index plots of the Pearson residuals and the Deviance residuals in Figure 3.3 indicated through visual inspection, two outliers, one close to observation no. 5000 and the other between no's 10000 and 15000. The index plot Leverage suggested that a case close to 14000 was an influential observation.

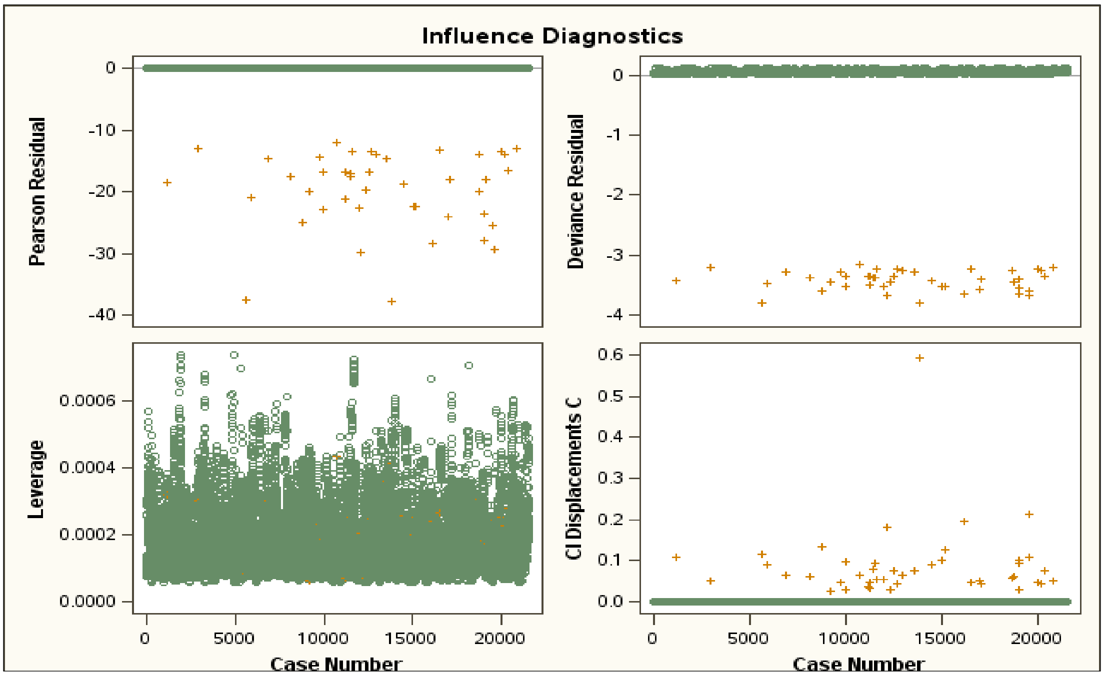


Figure 3.3: Residual and leverage plots from fitting the logistic model with 3 variables to the 48 months before default data.

The index plots of DFBETAS in Figure 3.4 indicated that a case close to no. 13000 caused instability in the parameter estimates for Cash Flow Ratio (m4).

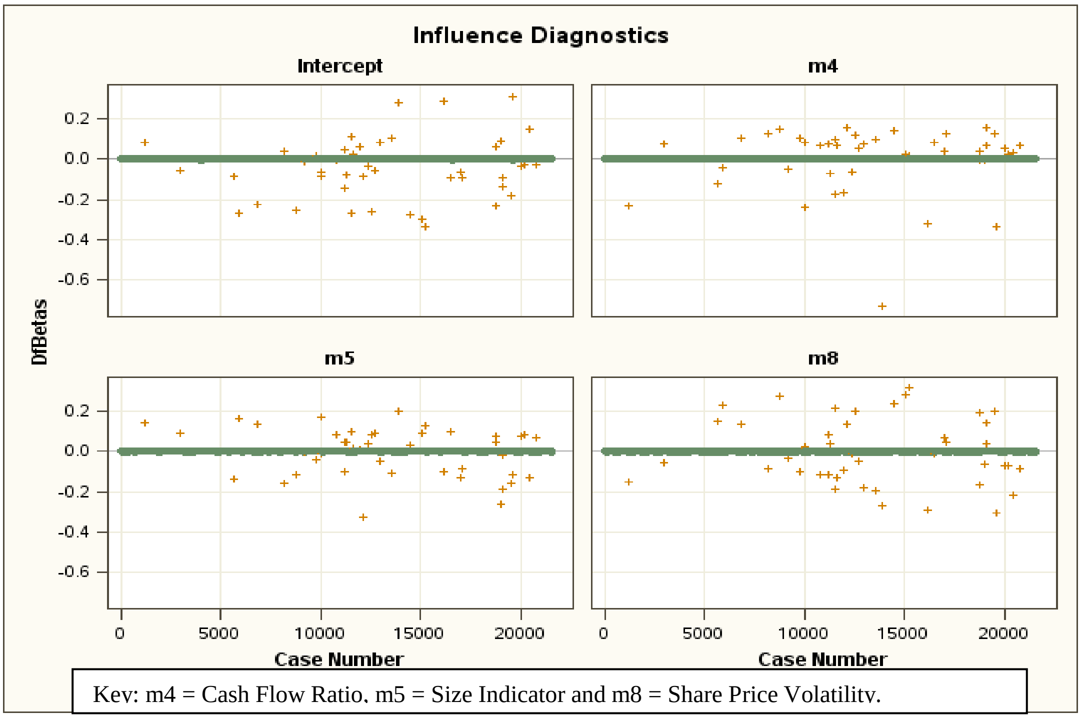


Figure 3.4: DFBETA plots from fitting the logistic model with 3 variables to the 48 months before default data.

It was once again important to note the fact that this was a very big data set (21520 observations), which means that this small number of outliers (one or two) wouldn't have much influence (if any) on the goodness of fitting the logistic regression model.

3.3.2 The fitted logistic regression model and inferences

In general terms, the outcome of the statistical inference normally gives a good indication of what should be done next in the model development process in terms of implementing the models as they are fit for purpose, or further experimentation. It is very important to note that during the statistical inferences to follow, from the various statistical techniques for logistic regression modelling (such as Goodness-of-fit tests, parameter significance tests, odds ratio analyses and ranking ability (ROC) tests), not only the statistical significance will be tested and reported, but also the practical significance by investigating the effect sizes. In Chapter 2 (Section 2.3), the measures of effect size associated with the various statistical techniques for logistic regression modelling have already been highlighted. It is also very important to remember that effect sizes should be used in conjunction with statistical significance testing as it gives additional insight and makes up for some of the "shortcomings" that exist in statistical significance testing as was discussed in Chapter 2 (Section 2.3.4).

3.3.2.1 Model diagnostics

The logistic regression model with the variables: Cash Flow Ratio (m4), Size Indicator (m5) and Share Price Volatility (m8) was fitted to the 48 month before default data using the Logistic Procedure in SAS Enterprise Guide Version 5.1. Table 3.20 contains the model fit statistics.

Table 3.20: Model fit statistics of the logistic model with 3 variables fitted to the 48 month before default data.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	634.854	613.132
SC	642.831	645.039
-2 Log L	632.854	605.132

No. of observations = 21520

A comparison of the model fit statistics of the model with an intercept only with those of the model with an intercept and the 3 variables: Cash Flow Ratio (m4), Size Indicator (m5) and Share Price Volatility (m8) shows that the model with an intercept and 3 variables fits better as the AIC was smaller for this model.

The p-values of the Wald, Score and Likelihood Ratio test and the null hypothesis that all 3 variables are insignificant in the model are displayed in Table 3.21

Table 3.21: Overall goodness-of-fit of the logistic model with 3 variables fitted to the 48 month before default data.

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	27.7226	3	<.0001
Score	24.6534	3	<.0001
Wald	21.7124	3	<.0001

The p-values are all less than 0.0001. Hence the null hypothesis is rejected and the model with 3 variables fitted significantly better than the empty model. This conclusion is supported by the Deviance and Pearson goodness-of-fit statistics in Table 3.22 where, for a Pearson Chi-square and Deviance statistic, a small p-value (<0.05) suggests that the fitted model is not adequate. What should however be taken into account is that a large difference between the Pearson statistic and the Deviance as was the case in Table 3.22 where the difference was almost 16000, provide evidence that the data are too sparse to use either statistic and the p-values are not valid and can be ignored (p=1).

Table 3.22: Deviance and Pearson Goodness-of-fit test statistics of the logistic model with 3 variables fitted to the 48 month before default data.

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	**Value**	**DF**	**Value/DF**	**Pr > ChiSq**
Deviance	576.9004	1.90E+04	0.0309	1
Pearson	16440.88	1.90E+04	0.8812	1

3.3.2.2 Goodness of fit of the model continued

The goodness of fit of the 48 month before default logistic regression model that was fitted was now checked using formal statistical tests (see Section 2.1.2.1). Table 3.23 displays the descriptive and test statistics of the models. Although there is no close analogous statistic in logistic regression to the coefficient of determination (R^2), the model summary Table 3.23 provides some approximations. As discussed in Chapter 2 (Section 2.1.2.1), Cox and Snell's R^2 attempts to imitate multiple R^2 based on 'likelihood', but its maximum usually is less than 1.0, making it difficult to interpret. Here it is indicating for the 48 month before default data sets, that only 0.1% of the variation in the "Indicator" variable is explained by the logistic model. The Nagelkerke modification that does range from 0 to 1 is a more reliable measure of the variation (similar to the proportion of the variance explained by the fitted model for linear regression R^2). Nagelkerke's R^2 will normally be higher than the Cox and Snell measure. In Table 3.23 it is 4% for the 48 month before default model, indicating a very weak relationship between the predictor variables and the predicted or indicator variable. These two tests do not really render the meaning of variance explained, as reported by Menard (2002) and do not correspond to predictive efficiency or can be tested in an inferential framework. For these reasons, a researcher can treat these two R^2 indices as supplementary to other, more useful evaluative indices, such as the overall evaluation of the model (like the ROC), tests of individual regression coefficients, and the Hosmer and Lemeshow goodness-of-fit test statistic. Field (2009) suggest that one should only consider the rough magnitude of these R^2 values and that these R^2 values are normally lower than the typical high R^2 values one is used to in goodness-of-fit tests for linear regression models. A more robust alternative to model chi-square is the Hosmer and Lemeshow test (refer to Chapter 2, (Section 2.1.2.1)). As reported by Hosmer and Lemeshow (2000), this test divided the observations or cases into 10 ordered groups based on their estimated probabilities where those with estimated probabilities below 0.1 formed the first group up to those with probabilities of 0.9 to 1.0 for the last group. Each of these categories was then further divided into two groups for comparison based on the actual observed outcome variable (default and non-default). The expected frequencies for each of the cells are obtained from the model. The probability (p) value is calculated from the chi-square distribution which has 8 degrees of freedom. The Hosmer and Lemeshow Goodness-of-Fit Test (H-L GOF) tests the hypotheses as discussed in Section 2.1.2.1. If, by means of the test, one fails to reject the null hypothesis ($p>0.05$), it means that the model is a good fit. If the H-L GOF test statistic was greater than .05 (using the 5% significance level), as would be the case for well-fitting models, the null hypothesis that there was no difference between observed and model-predicted values could not be rejected, implying that the model's estimates fitted the data at an acceptable level. Note that well-fitting models show non-significance on the H-L GOF test and this non-significance indicates that the model prediction of default or non-default does not significantly differ from what was observed.

In Table 3.23, the Hosmer and Lemeshow statistic for the 48 month before default model had a significance of 0.139 which means that in the fitted model, it was not statistically significant and therefore the model was a good fit. The result of Hosmer and Lemeshow Chi-square test has p-value of >0.05, which indicated that the estimated PD is quite close to the observed default rates for the 48 month before default model.

Table 3.23: The descriptive and test statistics of the goodness-of-fit of the logistic model with 3 variables fitted to the 48 month before default data set

Hosmer and Lemeshow Test			
Step	**Chi-square**	**df**	**Sig.**
1	12.278	8	.139

Model Summary

Step	-2 Log likelihood	Cox & Snell R-Square	Nagelkerke R-Square
1	605.132[a]	.001	.044

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001; Sig.=p-value.

3.3.2.3 Testing the significance of individual model parameters

The significance of the individual 3 parameters of the logistic model fitted to the 48 month before default data set was tested using the Wald test discussed in Chapter 2 (Section 2.1.1). Table 3.24 displays the results of the Wald test. The table shows that all the parameters were significantly different from zero ($p<0.05$) except for Share Price Volatility (m8), $p=0.2123>0.05$).

Table 3.24: Estimated coefficients of the logistic model with 3 variables fitted to the 48 month before default data.

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	5.636	0.4761	140.1292	<.0001
m4	1	2.2757	0.8302	7.5136	0.0061
m5	1	1.4866	0.7087	4.4	0.0359
m8	1	-0.7293	0.5848	1.5555	0.2123

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility.

It was however decided to include the variable Share Price Volatility (m8) in the model as well, after consultation with credit experts in business, who feel that according to their experience of Share Price Volatility, it is a very strong predictor of the probability to default. It is good business practice to involve the analysts when building the PD models, as their expert input can sometimes give a view that no statistical model alone can accurately capture from the underlying data. It is however still important to assess whether these variables that were included additionally will impact the overall goodness-of-fit (which was analysed in Tables 3.20 and 3.21) and whether they may cause multi-collinearity (see Section 3.3.2)

Table 3.25: Odds ratio estimates from fitting the logistic model with 3 variables to the 48 month before default data.

Odds Ratio Estimates			
Effect	**Point Estimate ($Exp(\hat{\beta})$)**	**95% Wald Confidence Limits**	
m4	0.103	.020	0.523
m5	0.226	.056	0.907
m8	2.074	.659	6.524

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility.

In Table 3.25 above, the coefficients were given as odds ratios which can be interpreted as the multiplicative change in the odds ratio of defaulting per unit change in the predictor variable. Note that the change can lie outside the interval [0 ; 1]. This means that the factor by which the odds of defaulting indicator increased or decreased needs to be put in the same scale in order for the results to make practical sense. For example, for a unit increase of 0.1 in the Cash Flow Ratio (m4), the odds of not defaulting (versus defaulting) increased by a factor of 1.2556. In order to determine this factor, the following first had to be considered:

If $Exp(\hat{\beta})$= 0.103 with $(\hat{\beta})$= 2.276 (coefficient of Cash Flow Ratio (m4)) and 2.276 x 0.1 = 0.2276, then Exp(0.2276) = 1.2556 (which is a scaled point estimate)

3.3.2.4 AUC and ranking ability of the fitted model

One other measure of the fitted logistic regression models' overall goodness-of-fit and their ability of discriminating between good and bad companies or those likely and not likely to default is the AUC from the model ROC curves. It was mentioned before, that the AUC together with Cohen's d are effect sizes used to quantify predictive accuracy. By using these measures and determining the effect sizes, we can check whether the current model is a valid model and whether the comparison of the Gini-scores of the validation sample with those of the training sample indicates that certain variables should be reconsidered in the selection process. In terms of model validity, the ROC curves and corresponding areas under the curve measures were constructed. The measure which is important to take note of when looking at the association between the Predicted Probabilities and Observed Responses is "c", where "c" is equivalent to the well-known measure AUC of ROC which was discussed in detail in Chapter 2 (Section 2.1.4). "c" ranges from 0.5 to 1, where 0.5 corresponds to the model randomly predicting the response or ratings and a 1 corresponds to the model perfectly discriminating between good and bad ratings (0 and 1 responses).

Table 3.26: The association of predicted probabilities and observed responses from fitting the logistic model with 3 variables to the 48 month before default data.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	71.8	**Somers' D**	0.437
Percent Discordant	28.2	**Gamma**	0.437
Percent Tied	0	**Tau-a**	0.002
Pairs	944944	**c**	0.718

From Table 3.26 and Figure 3.5, the discriminatory strength of the model is 71.8%. This is an indication of good ranking ability between "good" and "bad" counterparties by the fitted model. As per the guidance from Table 2.1 in Chapter 2, and AUC ≥0.7 is considered to have a large or strong effect size. It can therefore be seen from Table 3.26 and Table 2.1 that the ability of this fitted model to rank well between "good" and "bad" counterparties is strong.

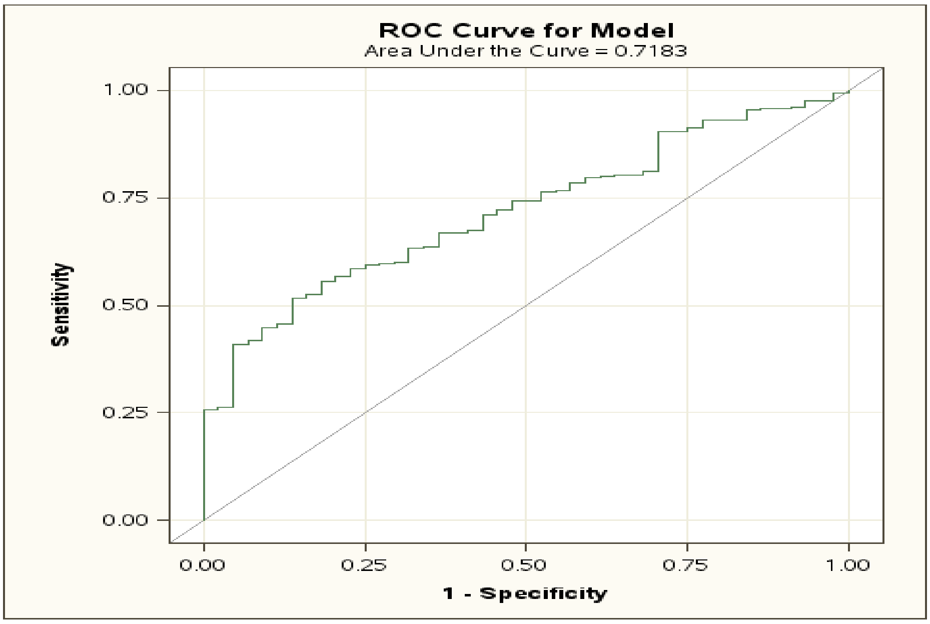


Figure 3.5: The ROC curve showing the area between the random and actual/fitted model from fitting the logistic model with 3 variables to the 48 month before default data.

3.4 Fitting the logistic regression models to the various months before default data sets

The process discussed in Section 3.1.4 can be repeated for all 48 data sets in order to select from the 9 variables, the final variables to be included in each of the 48 models, but for the purpose of explaining statistical and practical significance in even more detail, only the 1, 24 and 48 months before default data sets were analysed from here on further, as the same process applies to fitting all models and it would make this study too cumbersome if the analysis was performed for all 48 data sets. Through the logistic regression process (as just described in full detail for fitting the 48 month before default data to a model (refer to Section 3.3)) and additional judgemental or expert assessment, the following variables were included in the fitted models for the 1, 24 and 48 month before default data sets.

1 Month before default model

m1 = Excess Return on Share Price
m2 = Market Leverage
m3 = Gearing (Net Fair Value Debt / Net Tangible Assets)
m4 = Cash Flow Ratio (Cash Flow / Net Fair Value Debt)
m5 = Size Indicator (Minimum of Market Cap and Adjusted Turn Over);
m6 = Real Turnover Growth indicator;
m7 = Debtors / Turnover;
m8 = Share Price Volatility; and
m9 = (Three year high – Three Year Low)/ Share Price Three year low.

24 Months before default model

m1 = Excess Return on Share Price;
m4 = Cash Flow Ratio (Cash Flow / Net Fair Value Debt);
m5 = Size Indicator (Minimum of Market Cap and Adjusted Turn Over);
m6 = Real Turnover Growth indicator;
m8 = Share Price Volatility; and
m9 = (Three year high – Three Year Low)/ Share Price Three year low.

48 Months before default model

m4 = Cash Flow Ratio (Cash Flow / Net Fair Value Debt);
m5 = Size Indicator (Minimum of Market Cap and Adjusted Turn Over);
m8 = Share Price Volatility

Full model diagnostics, assumption testing, outlier analysis and stepwise selection have not been displayed for these models again (as it was discussed in detail for the 48 month before default model). Only the analysis and statistical tests that relate to effect sizes and practical significance was reported from here on further.

3.4.1 Correlation Analysis

Once the variables were selected, Pearson’s correlation matrices for the data sets being analysed were calculated (Table 3.27) and the selected variables in each data set were quickly assessed for possible high correlations and large effect sizes (using the conversion as per the guidance in Table 2.1 and as illustrated in Table 3.5). As can be seen from these three matrices, there were no variables included in the models that were very highly correlated (close to +/- 1.0 which could cause multi-collinearity issues in the fitted model) However, there were potential non-weak linear relationships between variables m2 and m3, and m2 and m4 in the 1 month before default data set, and between m8 and m9 in the 1 and 24 month(s) data sets (highlighted in red in Table 3.27).

Table 3.27: Pearson's correlation matrices (including final modelling variables) for 1, 24 and 48 months before default data sets.

1 month before default data set

	m1	m2	m3	m4	m5	m6	m7	m8	m9
m1	1	-0.251	-0.027	0.147	0.132	-0.069	-0.049	-0.294	-0.069
m2	-0.251	1	0.470	-0.521	-0.295	0.054	0.071	0.236	0.152
m3	-0.027	0.470	1	-0.363	-0.018	0.050	0.041	0.087	0.104
m4	0.147	-0.521	-0.363	1	0.255	-0.225	-0.162	-0.266	-0.214
m5	0.132	-0.295	-0.018	0.255	1	-0.240	-0.155	-0.358	-0.344
m6	-0.069	0.054	0.050	-0.225	-0.240	1	0.060	0.189	0.212
m7	-0.049	0.071	0.041	-0.162	-0.155	0.060	1	0.101	0.100
m8	-0.294	0.236	0.087	-0.266	-0.358	0.189	0.101	1	0.523
m9	-0.069	0.152	0.104	-0.214	-0.344	0.212	0.100	0.523	1

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low.

Pearson's r	r^2	Cohen's d	Effect	Variance in one variable accounted for by variance in other variable
-0.027	0.0007	-0.0540	Small	0%
0.236	0.0557	0.4857	Moderate	6%
-0.358	0.1282	-0.7668	Large	13%
-0.47	0.2209	-1.0650	Large	22%
-0.521	0.2714	-1.2208	Large	27%
0.523	0.2735	1.2272	Large	27%

24 month before default data set

	m1	m4	m5	m6	m8	m9
m1	1	0.130	0.098	-0.056	-0.281	-0.043
m4	0.130	1	0.227	-0.207	-0.256	-0.203
m5	0.098	0.227	1	-0.212	-0.348	-0.322
m6	-0.056	-0.207	-0.212	1	0.176	0.190
m8	-0.281	-0.256	-0.348	0.176	1	0.496
m9	-0.043	-0.203	-0.322	0.190	0.496	1

Key: m1 = Excess Return on Share Price, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m8 = Share Price Volatility; and m9 = Share Price high/low.

Pearson's r	r^2	Cohen's d	Effect	Variance in one variable accounted for by variance in other variable
-0.13	0.0169	-0.2622	Small	2%
-0.256	0.0655	-0.5296	Moderate	7%
0.496	0.2460	1.1424	Large	24%

48 month before default data set

	m4	m5	m8
m4	1	0.2047	-0.2392
m5	0.2047	1	-0.3300
m8	-0.2392	-0.3300	1

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility.

Pearson's *r*	r^2	Cohen's *d*	Effect	Variance in one variable accounted for by variance in other variable
0.2047	0.0419	0.4183	Small	4%
-0.33	0.1089	-0.6992	Moderate	11%

3.4.2 Goodness-of-fit of the models

The goodness of fit of each of the logistic regression models that were fitted was checked using descriptive statistics as well as using formal statistical tests (see Section 2.1.2.1). Table 3.28 displays the descriptive and test statistics of the models. Cox and Snell's R^2 is indicating respectively for the 1, 24 and 48 month data sets, that 0.8%, 0.2% and 0.1% of the variation in the "Indicator" variable is explained by the logistic model. Nagelkerke's R^2 will normally be higher than the Cox and Snell measure. In Table 3.28 it is respectively 28%, 8% and 4%, indicating rather weak relationships between the predictors and the prediction for the three models. As stated before (Section 3.3.4.2), these two tests does not really render the meaning of variance explained and does not corresponds to predictive efficiency or can be tested in an inferential framework. For these reasons, a researcher can treat these two R^2 indices as supplementary to other, more useful evaluative indices, such as the overall evaluation of the model (like the ROC), tests of individual regression coefficients, and the Hosmer and Lemeshow goodness-of-fit test statistic. If the H-L GOF test statistic was greater than .05 (using the 5% significance level), as would be the case for well-fitting models, the null hypothesis that there was no difference between observed and model-predicted values could not be rejected, implying that the model's estimates fitted the data at an acceptable level. In Table 3.28, the Hosmer and Lemeshow statistic respectively for the 1, 24 and 48 month data sets had a significance of 0.765, 0.793 and 0.139 which means that in all three models it was not statistically significant and therefore the models were quite a good fit. The result of Hosmer and Lemeshow Chi-square test has p-values>0.05, which indicated that the estimated PD are quite close to the observed default rates for the 1, 24 and 48 month before default models.

Table 3.28: The descriptive and test statistics for checking the goodness-of-fit of the logistic regression models fitted in Section 3.1.5

1 month before default model

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	4.927	8	.765

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	935.814[a]	.008	.279

a. Estimation terminated at iteration number 12 because parameter estimates changed by less than .001; Sig.=p-value.

24 months before default model

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	4.666	8	.793

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	893.787[a]	.002	.081

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001; Sig.=p-value.

48 months before default model

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	12.278	8	.139

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	605.132[a]	.001	.044

a. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001; Sig.=p-value.

3.4.3 Test of the significance of individual model parameters

The significance of the individual parameters of each of the models fitted in Section 3.1.5 was tested using the Wald test discussed in Chapter 2 (Section 2.1.1). Table 3.29 displays the results of the Wald tests as well as the confidence intervals for the odds ratios e^{β}, where $\hat{\beta}$ is the regression coefficient

Table 3.29: The Wald test statistics for testing the significance of the individual parameters of the fitted logistic regression models

1 month before default model

Variable	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	df	p-value	Exp($\hat{\beta}$)	Exp(($\hat{\beta}$)*0.1)	95% C.I.for EXP($\hat{\beta}$)	
								Lower	Upper
m1	-2.814	.755	13.888	1	.000	.060	0.7547	.014	.263
m2	2.377	.722	10.833	1	.001	10.777	1.2683	2.616	44.395
m3	.364	.404	.813	1	.367	1.439	1.0371	.652	3.174
m4	-3.693	1.268	8.485	1	.004	.025	0.6912	.002	.299
m5	-2.452	1.094	5.023	1	.025	.086	0.7825	.010	.735
m6	1.784	1.036	2.966	1	.085	5.956	1.1953	.782	45.386
m7	.933	.370	6.373	1	.012	2.543	1.0978	1.232	5.250
m8	2.917	1.019	8.203	1	.004	18.492	1.3387	2.512	136.150
m9	1.227	.709	2.990	1	.084	3.410	1.1305	.849	13.695
Constant	-10.160	1.088	87.157	1	.000	.000	0.3620		

Key: m1 = Excess Return on Share Price, m2 = Market Leverage, m3 = Gearing, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m7 = Debtors / Turnover, m8 = Share Price Volatility; and m9 = Share Price high/low; p-values>0.05 were highlighted; S.E.($\hat{\boldsymbol{\beta}}$) = Standard errors of the regression coefficients which can be used for hypothesis testing and constructing confidence intervals

24 months before default model

Variable	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	df	p-value	Exp($\hat{\beta}$)	Exp(($\hat{\beta}$)*0.1)	95% C.I.for EXP($\hat{\beta}$)	
								Lower	Upper
m1	.035	.397	.008	1	.930	1.036	1.0035	.475	2.256
m4	-1.167	.592	3.882	1	.049	.311	0.8899	.097	.994
m5	-1.618	.668	5.877	1	.015	.198	0.8506	.054	.733
m6	.569	1.332	.182	1	.669	1.766	1.0586	.130	24.018
m8	2.512	.683	13.543	1	.000	12.330	1.2856	3.235	46.987
m9	.721	.575	1.570	1	.210	2.056	1.0748	.666	6.349
Constant	-7.511	.651	133.255	1	.000	.001	0.4718		

Key: m1 = Excess Return on Share Price, m4 = Cash Flow Ratio, m5 = Size Indicator, m6 = Real Turnover Growth indicator, m8 = Share Price Volatility; and m9 = Share Price high/low; p-values>0.05 were highlighted; S.E.($\hat{\boldsymbol{\beta}}$) = Standard errors of the regression coefficients which can be used for hypothesis testing and constructing confidence intervals

48 months before default model

Variable	$\hat{\beta}$	S.E.($\hat{\beta}$)	Wald	df	p-value	Exp($\hat{\beta}$)	Exp(($\hat{\beta}$)*0.1)	95% C.I.for EXP($\hat{\beta}$)	
								Lower	Upper
m4	-2.276	.830	7.514	1	.006	.103	0.7964	.020	.523
m5	-1.487	.709	4.400	1	.036	.226	0.8618	.056	.907
m8	.729	.585	1.556	1	.212	2.074	1.0756	.659	6.524
Constant	-5.636	.476	140.128	1	.000	.004	0.5692		

Key: m4 = Cash Flow Ratio, m5 = Size Indicator and m8 = Share Price Volatility; p-values>0.05 were highlighted; S.E.($\hat{\beta}$) = Standard errors of the regression coefficients which can be used for hypothesis testing and constructing confidence intervals

Table 3.29 shows that:

- in the 1 month before default model all variables had significant effects (p-values<0.05) except Gearing (m3), Real Turnover Growth Indicator (m6) and Share Price high/low (m9) (p-values>0.05);
- in the 24 months before default model all variables had significant effects (p-values<0.05) except Excess Return on Share Price (m1), Real Turnover Growth Indicator (m6) and Share Price high/low (m9) (p-values>0.05); and
- in the 48 month before default model all variables had significant effects (p-values<0.05) except for Share Price Volatility (m8) (p-value>0.05).

The shortcomings of the tests in this section and their conclusions are that the conclusions are vague in the sense that nothing is said about how significant/insignificant the variable effects are (practical significance/insignificance). As examples the statistical as well as practical significance/insignificance of the Excess Return on Share Price (m1) and Cash-flow Ratio variable (m4) in the fitted 24 months before default logistic regression model were considered. From Table 3.29 it can be seen that the effect of Excess Return on Share Price was statistically insignificant (p-value=0.930>0.05). It basically means that the coefficient of this variable did not differ significantly from zero. On the other hand, the effect of the Cash-Flow Ratio was statistically significant (p-value=0.049<0.05), which means that the coefficient of this variable differed significantly from zero. One can however not say how significant this difference is. Therefore, to put this in terms of practical significance (or insignificance), the meaning and direction of the coefficients of Excess Return on Share Price and Cash-flow Ratio was interpreted. This was done in terms of the logged odds. The coefficient of 0.035 showed that a 0.1 unit increase (scaled to financial ratios) in the Excess Return on Share Price increased (because of + sign) the logged odds of defaulting by $0.035 \times 0.1 = 0.0035$ and the coefficient of -1.167 showed that a 0.1 unit increase in the Cash-flow Ratio lowered or decreased (because of – sign) the logged odds of defaulting by $1.167 \times 0.1 = 0.1167$. Further, the coefficients of these two variables were also translated into its effects on the odds of defaulting. In Table 3.29 the exponentiated value $e^{(\beta * 0.1)}$ of the coefficients were given as 1.0035 for the coefficient of Excess Return on Share Price and as 0.8899 for the coefficient of Cash-flow Ratio. Subtracting 1 from each of these exponentials of the coefficients and multiplying by 100 shows the % change in the odds of defaulting for a 0.1 unit change in each the variables as reported by Pampel (2000). A 0.1 unit increase in Excess Return on Share Price increased the odds of default by a multiple of 1.0035, or by 0.35%

(since (1.0035-1)×100=0.35). A 0.1 unit increase in Cash-flow Ratio decreased the odds of default by a multiple of 0.8899, or by 11.01% (since (0.8899-1) ×100 = -11.01).

The odds ratio of 1.036 for the coefficient of Excess Return on Share Price was a point estimate. The 95% confidence interval for the ratio was (.475;2.256) which included 1 and hence confirmed the insignificance of the effect of Excess Return on Share Price on the odds or log odds of default at the 0.05 level of significance (refer to Section 2.1.1). The odds ratio 0.311 for the coefficient of Cash-flow Ratio had a 95% confidence interval for the ratio of (0.097;0.994), which in this case, did not include 1 and hence confirmed the significance of the effect of Cash-flow Ratio on the odds or log odds of default at the 0.05 level of significance. When interpreting these results from a practical reporting point of view, the scaling factor of a 0.1 unit increase in Cash-flow ratio was again used. This means the 95% confidence interval of a percentage decrease in the odds of a company defaulting per increase in a company's cash-flow of one measuring unit is:

$$((1-(0.994\times0.1))\times100\%=90.06\% \; ; \; (1-(0.097\times0.1))\times100=99.03\%).$$

This means that one could say with a 95% certainty that the decrease in the odds of the probability of default (because of negative sign) lies between 0.9006 or 90.06% and 0.9903 or 99.03% if there's a 0.1 unit increase in the Cash-flow ratio. This interpretation of odds and log odds ratios and its confidence intervals make much more practical sense in the context of the business and its decision makers than just reporting their statistical significance as the effect of changes in the independent variables on the probability of default can be more easily interpreted.

3.4.4 Odds ratios as effect sizes

3.4.4.1 Goodness-of-fit check by means of the overall odds ratio (OOR)

The OOR which is also an effect size was used as a measure of goodness-of–fit of the fitted logistic regression models as well, as was discussed in Chapter 2 (Section 2.1.2.1). The Exp($\hat{\beta}$) values in Table 3.29 obtained the following overall odds values for the fitted logistic regression model.

- 1 month before default model: OOR=0.060×10.777×…×3.4410=1.9107
- 24 months before default model: OOR=1.036×0.311×…×2.056 =2.8610
- 48 months before default model: OOR=0.103×0.226×2.074 =0.0483

The respective Cohen's *d*'s for the 1, 24 and 48 month(s) before default models were (see Section 2.3.3):

$$\frac{\ln 1.9107}{1.81} = 0.3577, \frac{\ln 2.8610}{1.81} = 0.5808 \text{ and } \frac{\ln 0.0483}{1.81} = -1.6742 .$$

Thus, according to Table 2.1 the respective effect sizes for the fitted 1, 24 and 48 month(s) before default regression models are moderate/medium for the 1 and 24 month(s) before default models and large for the 48 month before default model. This means that the 48 months before default model predicted the odds of the probability of default better than the 1 and 24 month(s) before default models and is therefore of the three, the best credit risk model. Remember that the OOR of these models can be compare on the same scale as they were all converted to Cohen's *d*'s or effect sizes. It should be noted that the practical significance for the OOR can't really be reported the same way as odds ratios for the individual variables as there is not 1 measuring unit that is applicable for all the variables combined (like in the earlier example for the unit increase in the Cash-flow ratio and the odds of a decrease in the PD). It is basically just an index of how good the model is. The real advantage of having the OOR and the magnitude of it as an effect size, lies in the fact that now there is a standard unit of goodness-of-fit in Cohen's *d* which can be used to compare the goodness-of-fit of similar models. This ability to compare overall effects of models makes it practical and easier to report. It is almost like describing the characteristics of two entities as similar to apples and then being able to compare the apples with one another!

3.4.4.2 Significance of individual model parameters

One can also report the significance/insignificance of the effects of the variables on the odds or odds ratios of default in terms of the magnitude of the effect sizes of the odds ratios after converting them to Cohen's *d*. Again as examples, the variables Excess Return on Share Price (m1) and Cash-flow Ratio variable (m4) in the fitted 24 months before default logistic regression model were considered. The respective odds ratios of Excess Return on Share Price (m1) and Cash-flow Ratio (m4) (from Table 3.29) were 1.036 and .311, and the respective Cohen's *d*'s are:

$$\frac{\ln 1.036}{1.81} = 0.0195 \text{ and } \frac{\ln 0.311}{1.81} = -0.6443.$$

Thus, according to Table 2.1 the respective effect sizes of variables Excess Return on Share Price and Cash-flow Ratio were small/weak and medium/moderate respectively. These tie back to the effect of Excess Return on Share Price being statistically insignificant to the model while the effect of Cash-flow Ratio was statistically significant

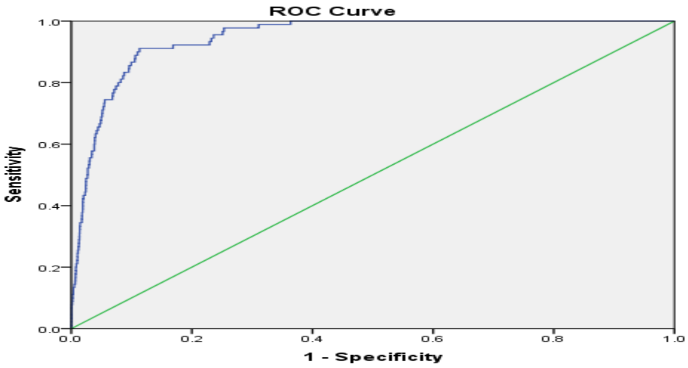
This whole process of reporting statistical significance versus practical significance can be repeated for each coefficient in each of the models. This will facilitate better business understanding and interpretation about the relationship and contribution of the different variables (financial ratios) to the likelihood of companies to default.

3.4.4.3 AUC and ranking ability of the fitted logistic models

One other measure of the fitted logistic regression models' overall goodness-of-fit and their ability of discriminating between good and bad companies or those likely and not likely to default is the AUC from the model ROC curves (refer to Section 3.3.4.4).

It was already mentioned before, that the AUC together with Cohen's *d* is an effect size used to quantify predictive accuracy. By using these tests and determining the effect sizes, we can check whether the current model is a valid model and whether the comparison of the Gini-scores of the validation sample to the training sample indicates that certain variables should be reconsidered in the selection process. In terms of model validity, the ROC curves and corresponding areas under the curve measures were constructed.

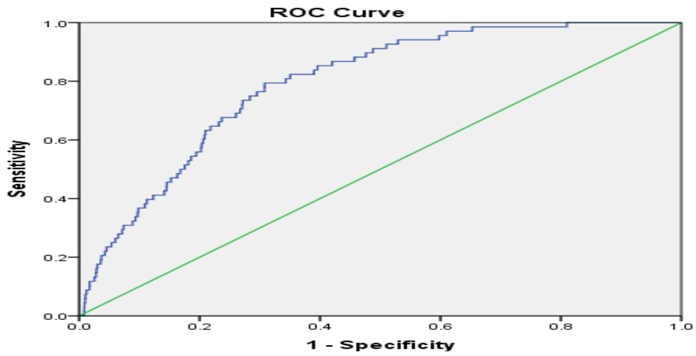
1 month before default model



AUC	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.947	.008	.000	.932	.961

Figure 3.6 (a) and Table 3.30 (a): The ROC Curve and corresponding AUC of the fitted logistic regression model.

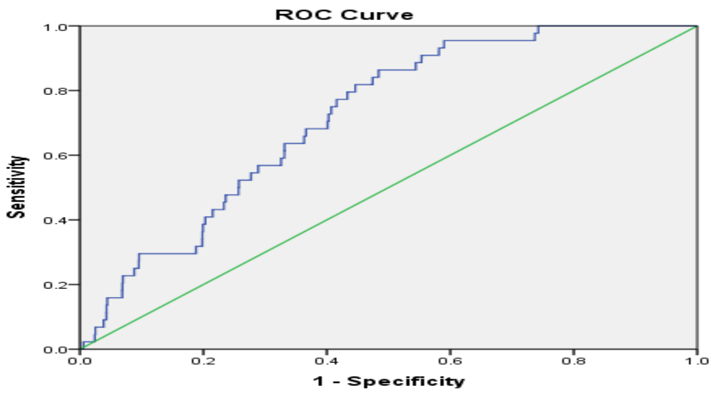
24 months before default model



AUC	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.793	.022	.000	.750	.836

Figure 3.6 (b) and Table 3.30 (b): The ROC Curve and corresponding AUC of the fitted logistic regression model.

48 month before default model



AUC	Std. Errora	Asymptotic Sig.b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.718	.030	.000	.660	.776

Figure 3.6 (c) and Table 3.30 (c): The ROC Curve and corresponding AUC of the fitted logistic regression model.

In Figure 3.6 and Table 3.30, the AUC for the 1 month before default model is 0.947 which means that the model's ranking ability of 94.7% is close to 100% which is an indication of perfect prediction and far from 50%, which would have indicated random prediction. This model therefore has a very high discriminatory power. Correspondingly, the magnitude of the effect size Chapter 2, (Section 2.3.3) for AUC is > 0.7 and therefore large. This indication of a high ranking ability is reporting of practical significance as it is easy for decision makers to interpret the significance of the model in the context of its purpose.

Similarly, the 24 Month and 48 Month models which gives ranking abilities of 79.3% and 71.8% can also be seen as having large effect sizes (AUC > 0.7) as per the guidance from Table 2.1 in Chapter 2, which means that in terms of practical significance, they both have the ability to rank well between good and bad companies and subsequently predict default accurately.

In order to test whether the models will still perform well after a year or two (out of time), the models are fitted or trained by using the training data sets and assessed by applying the model to the validation data set. This out of time performance tests were done first with 2008 and 2009 and then with 2008 and 2010 data. From the results, comparisons can be done to determine whether the predictability and ultimately the performance of the models over time improved, deteriorated or stayed stable. For the models which were tested, the ranking remained stable and the models are deemed to be valid over time, as can be seen below in Table 3.31.

Table 3.31: Ranking ability of models over time (AUC Comparisons), which is an indication of the stability in predicting default for the fitted logistic regression models. (Only Months 1, 24 and 48 are shown)

	1 Month before Default	24 Months before Default	48 Months before Default
2009	93.92%	80.21%	73.52%
2008	94.66%	79.29%	71.83%
%Change in AUC	-0.74%	0.92%	1.69%
2010	94.19%	79.76%	72.83%
2008	94.66%	79.29%	71.83%
% Change in AUC	-0.47%	0.47%	1.00%

The process of determining AUC, reporting on its practical significance and comparing training and validation data sets can be repeated for all 48 models as part of a validity test.

3.5 Constructing the PD models

Once all 48 logistic regression models were fitted and the assumptions, goodness-of-fit and significance (both statistical and practical) testing were done by following the process described in Sections 3.3 and 3.4, the final PD model could be created. As an example to explain the process of rating a counterparty, ABC Ltd, a corporate company listed on the JSE (which is a fictitious company used for the purpose of this demonstration) was rated by calculating its final PD. The process of creating the final PD model, as discussed in Chapter 2 (Section 2.2) entails the following. First the coefficients of all 48 logistic regression models were considered, where the coefficients α_0, β_1, β_2,.....,β_k (where k=9) were the regression coefficients estimated from the data by the maximum likelihood method for each model (refer to Chapter 2 (Section 2.1)). The coefficients for these models were summarised in Table 3.32 below. This table also gives a clear indication of which variables were ultimately included into each of the 48 models. For the 1 month before default it can for example be seen that all 9 variables were included, while for the 48 month before default model only Cash Flow Ratio (m4), Size Indicator (m5) and Share Price Volatility (m8) were included in the final model.

Table 3.32: Coefficients of the 48 fitted logistic regression models

Months before default	Constant	Excess Return (m1)	Market Leverage (m2)	Gearing (m3)	Cash Flow Ratio (m4)	Size indicat or (m5)	Real Turn Over Growth (m6)	Debtors/ Turn Over (m7)	Share Price Volatility (m8)	Share Price High Low (m9)
1	-10.16	-2.81	2.38	0.36	-3.69	-2.45	1.78	0.93	2.92	1.23
2	-10.35	-2.32	2.50	0.28	-3.09	-2.35	1.37	1.13	3.25	0.86
3	-9.64	-2.15	2.16	0.29	-2.95	-2.05	1.06	1.12	2.69	0.93
4	-10.09	-1.64	2.49	0.20	-2.44	-1.81	1.01	1.22	2.75	0.92
5	-9.32	-1.60	1.83	0.23	-2.74	-1.68	0.66	1.16	2.38	1.13
6	-8.99	-1.91	1.59	0.31	-2.30	-1.59	1.29	1.29	2.01	1.11
7	-8.43	-2.07	1.05	0.54	-2.62	-1.74	0.77	1.40	1.64	1.22
8	-7.92	-2.57	0.81	0.57	-2.24	-1.61	0.55	1.38	1.39	1.18
9	-8.36	-1.84	0.76	0.51	-2.13	-1.43	0.59	1.28	1.67	1.38
10	-7.97	-1.99	0.34	0.67	-2.05	-1.43	0.37	1.25	1.70	1.22
11	-8.04	-2.03	0.50	0.36	-2.03	-1.21	0.75	1.25	1.68	1.37
12	-8.03	-2.25	-0.13	0.48	-1.83	-1.24	1.05	1.67	1.70	1.43
13	-6.90	-2.17		-	-1.77	-1.37	0.87	-	1.96	1.30
14	-6.98	-1.67	-	-	-2.16	-1.58	1.00	-	2.13	1.18
15	-6.89	-1.17	-	-	-2.49	-1.56	1.01	-	1.98	1.10
16	-6.80	-1.10	-	-	-2.27	-1.65	0.82	-	1.91	1.07
17	-6.83	-1.20	-	-	-1.71	-1.71	0.95	-	1.83	1.17
18	-6.95	-1.00	-	-	-1.62	-1.71	0.58	-	1.98	1.16
19	-6.80	-1.16	-	-	-1.61	-1.61	0.57	-	1.93	1.05
20	-6.87	-0.73	-	-	-1.97	-1.52	0.50	-	1.83	1.15
21	-7.01	-0.61	-	-	-2.19	-1.29	0.42	-	1.75	1.34
22	-7.23	-0.30	-	-	-2.23	-1.25	0.78	-	2.05	1.14
23	-7.13	-0.18	-	-	-2.21	-1.36	0.33	-	1.97	1.12
24	-7.51	0.04	-	-	-1.17	-1.62	0.57	-	2.51	0.72
25	-7.76	0.20	-	-	-1.39	-1.41	-	-	2.64	0.91
26	-7.91	0.11	-	-	-1.06	-1.48	-	-	2.87	0.90
27	-7.96	0.22	-	-	-0.88	-1.54	-	-	2.78	0.98
28	-7.68	0.35	-	-	-1.16	-1.58	-	-	2.30	1.11
29	-7.45	-0.02	-	-	-1.22	-1.40	-	-	1.65	1.62
30	-7.62	-0.12	-	-	-1.15	-1.22	-	-	1.70	1.80
31	-7.55	-0.16	-	-	-0.98	-1.40	-	-	1.89	1.50
32	-7.68	0.06	-	-	-1.01	-1.34	-	-	2.03	1.44
33	-7.66	0.23	-	-	-1.01	-1.38	-	-	1.97	1.39
34	-7.93	0.50	-	-	-0.93	-1.30	-	-	2.46	1.07
35	-7.41	0.37	-	-	-1.07	-1.43	-	-	1.96	1.03
36	-7.07	0.04	-	-	-1.40	-1.45	-	-	1.85	0.94
37	-6.86	-0.09	-	-	-1.28	-1.44	-	-	1.76	0.81
38	-6.60	0.10	-	-	-1.30	-1.60	-	-	1.58	0.52
39	-6.36	-	-	-	-1.34	-1.63	-	-	1.79	-
40	-6.49	-	-	-	-1.15	-1.61	-	-	1.92	-
41	-6.19	-	-	-	-1.35	-1.48	-	-	1.49	-
42	-6.04	-	-	-	-1.16	-1.68	-	-	1.23	-
43	-5.85	-	-	-	-1.29	-1.79	-	-	1.03	-
44	-5.99	-	-	-	-1.29	-1.71	-	-	1.25	-
45	-5.85	-	-	-	-1.58	-1.65	-	-	1.07	-
46	-5.94	-	-	-	-1.64	-1.64	-	-	1.19	-
47	-5.82	-	-	-	-1.44	-1.63	-	-	0.93	-
48	-5.64	-	-	-	-2.28	-1.49	-	-	0.73	-

From the coefficients in Table 3.32, the PD_i, which is the probability of default obtained from the fitted i^{th} month before default logistic regression model (where *i=1,2,…,48 and k=1,2,…,9*) was constructed by using the formula:

$$PD_i = \frac{1}{1+\exp[-(\hat{\alpha}_{i0}+\hat{\beta}_{i1}m_1+\hat{\beta}_{i2}m_2+...+\hat{\beta}_{ik}m_k)]}$$

For ABC Ltd, the following financial ratios were obtained from its monthly financial statements for January 2010.

Table 3.33: Financial Ratios for ABC Ltd

Variable Name		Ratio
ExcessReturn	m1	0.9
Market Leverage	m2	0.05
Gearing	m3	0.09
CashFlow Ratio	m4	0.9553
Size Indicator	m5	0.8832
realTurnOver Growth	m6	0.00
Debtor/Turnover	m7	0.13
Share Price	m8	0.51
Share_High_Low	m9	0.65

By using these 9 ratios and substituting them into the 48 equations for the 48 models, the following PDs were obtained:

Table 3.34: PDs obtained from the 48 months before default logistic regression models

Month	PD		Month	PD		Month	PD
1	0.000000		17	0.000081		33	0.000491
2	0.000000		18	0.000100		34	0.000577
3	0.000001		19	0.000101		35	0.000504
4	0.000002		20	0.000107		36	0.000336
5	0.000003		21	0.000117		37	0.000363
6	0.000004		22	0.000125		38	0.000366
7	0.000004		23	0.000136		39	0.000323
8	0.000005		24	0.000263		40	0.000374
9	0.000012		25	0.000287		41	0.000379
10	0.000014		26	0.000325		42	0.000400
11	0.000017		27	0.000390		43	0.000350
12	0.000018		28	0.000362		44	0.000361
13	0.000048		29	0.000372		45	0.000312
14	0.000040		30	0.000415		46	0.000298
15	0.000045		31	0.000402		47	0.000367
16	0.000056		32	0.000456		48	0.000211

Then for *i=2,3,…,48*, the cumulative probability of default for month *i* is given by:

$$CPD_i = PD_1 + \sum_{l=2}^{i} PD_l \prod_{j=1}^{l-1} (1 - PD_j)$$

And by applying this formula to the 48 PDs as obtained in Table 3.34, 48 cumulative PDs were calculated:

Table 3.35: Cumulative PDs for the 48 models

Month	PD		Month	PD		Month	PD
1	0.000000		17	0.000350		33	0.004788
2	0.000000		18	0.000449		34	0.005362
3	0.000001		19	0.000550		35	0.005863
4	0.000003		20	0.000657		36	0.006197
5	0.000006		21	0.000774		37	0.006558
6	0.000010		22	0.000899		38	0.006922
7	0.000014		23	0.001035		39	0.007243
8	0.000019		24	0.001297		40	0.007614
9	0.000031		25	0.001584		41	0.007990
10	0.000046		26	0.001908		42	0.008387
11	0.000063		27	0.002298		43	0.008734
12	0.000081		28	0.002659		44	0.009092
13	0.000129		29	0.003030		45	0.009401
14	0.000168		30	0.003444		46	0.009697
15	0.000213		31	0.003845		47	0.010060
16	0.000269		32	0.004299		48	0.010269

The cumulative default probabilities (CPD_ms) for months 12, 24, 36 and 48 (years 1, 2, 3 and 4) are annualized with the following formula:

$$AnnPD_N = 1 - (1 - CPD_N)^{\frac{12}{N}}, \; N = 12, 24, 36, 48$$

When applying this formula to the 48 cumulative PDs, 4 annualised PDs for years 1, 2, 3 and 4 were obtained.

Table 3.36 Annualised PDs from the 48 cumulative PDs

Month	Annualized PD
12 (Year 1)	0.000081
24 (Year 2)	0.000649
36 (Year 3)	0.002070
48 (Year 4)	0.002577

The final long-run average one year probability of default (*FinalPD*) that is assigned to ABC Ltd. is the maximum of the 12, 24, 36 and 48 month annualised probabilities. The reason why the maximum of the $AnnPD_N s$ were taken was to ensure conservatism in assigning a rating to the counterparty.

The formula used was:

$$FinalPD = Max(AnnPD_{12}, AnnPD_{24}, AnnPD_{36}, AnnPD_{48})$$

$$= \text{Max}\ (0.000081, 0.000649, 0.002070, 0.002577)$$

$$= 0.0026$$

Therefore, the probability that ABC Ltd. will default is 0.26% (based on its current financial statements). This PD rating will then be used by the bank to determine what the counterparty's risk profile looks like, how big the loan is that will be made to the client, at what rate the money is borrowed to the client and how much capital reserves the bank must hold against this client. The details of this process falls outside the ambit of this thesis and therefore the paper was limited to the construction and testing of, and reporting on the underlying data and fitted logistic regression models used to construct the PDs.

4 Conclusion, recommendations and applications

4.1 Statistical and practical significance revisited

The objective of this study was to investigate the use of effect sizes as statistical indices for practical significance in predictive models used in the financial credit risk environment. The more specific objectives of this thesis were to take a logistic regression model used to predict probability of default and use practical significance to perform the following:

1. Correlation tests, where the conversion of Pearson's correlation coefficients to Cohen's *d* values and the corresponding indication of the magnitude of the effect sizes assisted in deciding which variables to include, exclude or maybe further investigate through techniques like Cluster Analysis for correlation.
2. Variable selection tests, where not only statistical significance testing was performed which would give an indication of the significance of the contribution of a parameter to the overall probability of default, but multiplicative changes in the odds ratio of defaulting per unit change in the predictors was expressed as well.
3. Risk measures - Estimation of risk or performing of risk tests (assess the ability to discriminate between high and low risk), where the area under the ROC was demonstrated and utilised as a measure to indicate how well the rating model can distinguish between or rank low and high risk counterparties and therefore, how valid the model is in predicting default.
4. Goodness of fit tests, where, Pseudo R^2-values were interpreted as measures of how much of the variation in the "Indicator" variable is explained by the logistic model and overall odds ratios. The advantage of having the overall odds ratio and the magnitude of it as an effect size for logistic models as well, was displayed in the fact that now there is a standard unit of goodness-of-fit in Cohen's d which can be used to compare the goodness-of-fit of similar models.

First, the principles of credit risk modelling were unpacked. Once this was done, credit risk models were developed. After all the underlying assumptions were met, logistic regression models were fitted and used, as these models have the ability to predict a binary outcome, which in this case was to predict whether a counterparty will be able to repay his loan or not. In order to assess whether these models are fit for this purpose, the traditional goodness-of-fit and significance testing were performed. The different aspects of model development and testing were therefore evaluated, by not just looking at statistical significance, but also in the context of effect sizes and practical significance. This included the investigation of strength of certain relationships, obtaining goodness-of-fit measures which were comparable between different models and looking at the impact of change in unit values in predictor variables on the predicted variable. This was done by determining effect sizes, explaining what the magnitude of these effect sizes mean and how they could be converted to a standardised (comparable) value (Cohen's *d*). Most importantly, this study displayed not only how significance could be reported from a statistical point of view, but also in a much more practical way that won't only make sense to the statistician or model builder who is normally interested in whether or not to reject a null-hypothesis, but to a much wider audience. In the environment where these statistical models are implemented, they are not standalone prediction tools anymore, but they form part of a decision-making process where their results need to be understood and interpreted in the environment they are operating in. Therefore, interpretation of model testing and results as well as the reporting thereof needs to extend beyond hypothesis testing, p-values, R^2s, AUC and correlation coefficients, etc. In this study it is shown how results could be reported in such a way that business decisions can be made based on these strengths, ratios and relationships. These effect sizes and interpretation thereof not only helps to improve the modelling process, but also the understanding of the results coming from the analysis in terms of the context in which these models will be used. Practical significance is all about contextualising how one reports statistical results. Another feature of effect size that was demonstrated in this study was that it could be directly converted into statements about a comparison of two measures like the AUC for models fitted to different data sets.

4.2 Limitations of the model

As the model was developed on data that spanned over a certain time period (January 1995 – May 2008), the model may have to be re-calibrated to more recent data before being used to rate a counterparty that has more recent data available. Re-calibration refers to the process of fitting more recent data to the model, assessing the stability of the input parameters and testing the goodness-of-fit of the model when fitting the model to the bigger data set. The model and its performance are also dependent on economic cycles, as explained in Chapter 1 (Section 1.5.1). This means that during a recession or "downturn" period the model may react different than during a period of financial and economic growth and therefore, the most recent of both of these periods need to be included in the data in the model. Another data limitation speaks to the fact that data from companies that are listed with the JSE was used to develop this model, which means that the model is restricted to South African specific data and cannot be fitted to international data or data from other countries. Therefore, results from the model cannot be generalised to, for example, the American market or the model cannot be applied to a counterpart that is listed on, for example, the New York Stock Exchange. In terms of model limitations, only random effects were considered in constructing the link function, while fixed effects like sector or industry type, as can be seen in Table 3.2 was not considered. There may be "concentration" risk due to certain industry effects that may not currently be included into the model and therefore bias may exist in the model due to omission of a fixed factor like sector/industry type. If the effect of industry type needs to be assessed as part of rating counterparty, a mixed model may have to be developed and used (refer to Dietsch and Petey (2011)).

4.3 Advantages and disadvantages of the use of effect sizes

It is once again very important to note that; how well the logistic regression model fitted the data, whether this was the best model to fit and the disadvantages of using this model was not the key focus point of the thesis, but rather how the variable selection testing, model diagnostics and goodness-of-fit test results were **translated** in to practical terms (effect sizes) and reported on a scale or level that is understandable to all business stakeholders and decision makers. What is therefore very important is to highlight, is strengths and weaknesses as well as advantages and disadvantages that may exist in using effect sizes in credit rating models.

4.3.1 Advantages

An advantage of using standardised effect sizes like Cohen's d (as was described in Chapter 2 and demonstrated in Chapter 3), allowed for quantifying effects measured on different or arbitrary scales and for comparing the relative sizes of effects from different models fitted to different data sets. When standardized effect sizes are used, these estimates can be compared to reach a conclusion.

Another advantage of the effect size analyses that were performed in this study, was that it assisted in creating tables that displayed information in a coherent manner, therefore allowing easy comparisons against one another and pre-determined benchmarks (for example, one can look at the predetermined correlation threshold of $|r| > 0.4$ or Cohen's $d > 0.868$ which is an indication of high correlated or a large effect size).

4.3.2 Disadvantages

During this study, it was the experience of this researcher that most of the well-known statistical software packages tend to offer limited functionality for creating effect sizes and the conversion from statistical significance measures to effect sizes like Cohen's *d* still has to be done manually. The reporting of effect sizes has brought about a new school of thought in the statistical world and the caution that one needs to take in such a case, is to "over" use or emphasise this new approach and thereby underplay traditional statistical significance testing. Warning lights of this practice can already be seen in the social science environment, where since the 1980s there has been some key publications (refer to Wilkinson (2001)) to educate social scientists about the misuse of significance testing and the need for more common reporting of effect sizes. This publication went to the opposite extreme of statistical significance testing by stating that statistical research that doesn't report effect sizes is considered inferior. One should rather value an approach that includes complete reporting of statistical tests combined with descriptions of both the data and the effects. In the financial risk environment, the reporting of effect sizes is still in a much more infant phase than in the social sciences and therefore it is key to ensure that there is always a good balance and combination present in reporting statistical and practical significance.

A further potential disadvantage, as Cohen (1988) acknowledged and as can also be seen in this study, using terms like 'small/weak', 'medium/moderate' and 'large/strong' out of context can be dangerous and that is the reason why the effectiveness of a specific intervention can only be interpreted in relation to other interventions that seek to produce the same effect. Only in such a case can, for example, Cohen's description of an effect size 0.5 as medium/moderate is used in the same context throughout.

4.4 Practical significance in the context of credit risk rating models

It is very important to remember that practical significance is not drawing a line through statistical significance. It is rather the case as reported by Wilkinson and APA (1999) that it is almost always necessary when reporting practical significance, to include some index of effect size or strength of relationship in the results section. The general principle is to provide the reader, person or committee reported to, not only with information about statistical significance but also with enough information to assess the magnitude of the observed effect or relationship. Therefore statistical and practical significance together can contribute to more sensible reporting

To relate this to the specific situation in the bank, as a way forward, the following is suggested as an application of practical significance: Any model that is developed to be used to predict default must currently go through a rigorous governance process (a requirement from the national regulator which is the SARB). This includes peer reviews, independent validations of the models and presenting the results and performance of the models to a technical governance committee for approval before the models are implemented. The nature of reporting the results and performance is very statistical and makes a lot of sense to Statisticians, Mathematicians, Actuaries and Financial Engineers, but is sometimes not that easy to interpret for Accountants, Business Analysts, Credit Officers and Auditors. Unfortunately the technical governance committees as mentioned above (with the responsibility of challenging, interrogating and approving statistical models) are made up by people with skills ranging from quantitative to business or legal and qualifications ranging from Statisticians to Auditors. It is sometimes even necessary to report on the models in non-technical credit, policy and risk appetite committees. It is therefore recommended that effect sizes is determined and interpreted as part of model development and statistical analysis that is presented to these committees for model approval. This will enable the reporting of statistical as well as practical significance. If the practical significance of the models is reported in the business and strategy forums, it will align decision making more to what the models try to predict and it will help the credit and business experts to get more involved in the model development as they will better understand the underlying assumptions, relationships and variables in the models as well as what the models try to achieve. As a practical example of this, it will make much more sense to report to a credit analyst or accountant that needs to make decisions about product growth or business expansion and strategy, that a statistically significant model was used to determine that a 0.1 unit increase in Excess Return on Share Price increased the odds of default by 0.35%, than to report to them purely that the p-value of Excess Return on Share Price was <0.05 and therefore the coefficient was statistical significantly different from 0. A much more informed business decision can then be made based on what was reported practically.

4.5 Future research and the use of alternative models

In this study, a credit rating model that indicates the credit worthiness and probability of default of a credit counterparty was used to demonstrate how results can be measured in terms of statistical tests and effect sizes and how these results can be reported in terms of statistical significance and practical significance. This study can however be expanded to other applications in the modelling of credit risk as well. Examples of these are credit application and behavioural scorecards (refer to Siddiqi (2006)) and living expense models (refer to ISI (2013)). The application scorecards are used by financial institutions as automated statistical tools that approve, refer or decline consumers who apply for products such as home loans, vehicle finance, credit cards, etc. The behavioural scorecards are used by institutions to track the transactional behaviour and predict future behaviour of existing customers against predefined profile benchmarks. Living expense models are used by financial institutions to predict living expenses and disposable income of customers as part of assessing whether the customer can afford the loan or product applied for. The models/scorecards just mentioned are typically linear or logistic regression models, as they are used to predict a future outcome or behaviour and therefore, the use of effect sizes in the development, analysis and reporting of results are similar to what was demonstrated in the credit rating models demonstrated in this thesis. As reported by Siddiqi (2006), these scorecards drive the marketing and expansion strategies of most financial institutions. Marketing and strategy typically does not sit with model builders and quantitative teams, but rather with marketing and executive committees and therefore practical reporting of model results to these committees may give them a much better understanding of how the models can enhance their future strategies.

Although logistic regression models are used predominately in credit rating models, other statistical models that are used by financial institutions to make credit rating and scoring decisions are Decision Trees, Discriminant Function Analysis Models and Generalised Linear Mixed Models (GLMM). According to Middleton (2007), a decision tree can be used as a model for sequential decision problems under uncertainty. A decision tree describes graphically the decisions to be made, the events (default or non-default in the case of credit rating models) that may occur, and the outcomes associated with combinations of decisions and events. Probabilities are assigned to the events, and values are determined for each outcome. A major goal of the analysis is to determine the best decisions. According to Hair *et al*. (1998) discriminant function analysis is a statistical analysis to predict a categorical dependent variable (default or non-default) by one or more continuous or binary independent variables (predictor variables). Discriminant function analysis is a categorisation tool, but assumptions of normality and sensitivity to outliers need to be taken into consideration in these models. As already discussed in Section 4.2 above, the current model was developed by using random effects only, but fixed effects like sector type and time-periods can be considered as well. As reported by Dietsch and Petey (2011), a GLMM can be developed that takes into account both random and fixed effect. This GLMM would enable one to estimate the impact of additional risk factors like industry concentration on credit risk (i.e. the risk of defaulting). When analysing the statistical tests and results of these models, the statistical significance testing can also be converted to effect sizes and reported practically, similar to what was demonstrated in this thesis with the logistic regression model.

Bibliography

Acion L., Peterson J. J., Temple S. and Ardnt S., 2006. Probabilistic index: An intuitive non-parametric approach to measuring the size of treatment effects. *Statistics in Medicine, 25*.

Allen J. and Le H., 2008. An Additional Measure of Overall Effect Size for Logistic Regression Models. *Journal of Educational and Behavioural Statistics*, *33*.

Bangia A., Diebold F.X. and Schuermann T., 2000. Ratings Migration and the Business Cycle, With Applications to Credit Portfolio Stress Testing. Last accessed 26 June 2014. Available at: http://fic.wharton.upenn.edu/fic/papers/00/0026.pdf

Basel II Accord., 2005. Basel Committee on Banking Supervision - International Convergence of Capital Measurement and Capital Standards. *A Revised Framework.*

Chinn S., 2000. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19.*

Cohen J., 1977. Statistical Power Analysis for Behavioural Science. *2nd Edition.*

Cohen J., 1988. Statistical Power Analysis. *2nd Edition.*

Cox, D. R. and Snell E.J., 1989. The Analysis of Binary Data. *2nd Edition.*

Dietsch M. and Petey J., 2011. Sector concentration risk in SME credit portfolios: A multifactor approach.

Dunst C.J., Hamby D.W. and Trivette C.M., 2004. Guidelines for Calculating Effect Sizes for Practice Based Research Syntheses. *Centerscope*. *3*(1).

Engelmann B. and Raumeier R. 2006. The Basel II Risk Parameters: Estimation, Validation and Stress Testing. *2nd Edition.*

Field A., 2009. Discovering Statistics Using SPSS. *3rd Edition.*

Flach P., 2009. ROC Analysis: Encyclopaedia of Machine Learning

Goodyear M., 2011. Extending COPD Prevalence Modelling to Very Small Areas. Last accessed 12 June 2014. Available at: http://www.lewishamjsna.org.uk/Reports/Extending%20COPD%20Prevalence%20Modelling%20to%20Very%20Small%20Areas%20Internal%20report.pdf

Grissom Robert J. and Kim John J., 2005. Effect Sizes for Research: A Broad Practical Approach. *1st Edition.*

Gutierrez-Osuna R., 2011. Pattern Analysis: Cross Validation. Last accessed 28 May 2014. Available at: http://research.cs.tamu.edu/prism/lectures/pr/pr_l13.pdf

Hair J.F., Anderson R.E., Tatham R.L. and Black W.C., 1998. Multivariate Data Analysis. *5th Edition.*

Hosmer, D. and Lemeshow, S., 2000. Applied Logistic Regression. *2nd Edition.*

Hunter J. E. and Schmidt F. L., 2004. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. *2nd Edition.*

ISI., 2013. Insolvency Service of Ireland - Guidelines on a reasonable standard of living and reasonable living expenses.

Investopedia. “Financial Terms". Investopedia US. Last accessed 30 March 2015. Available at: http://www.investopedia.com/dictionary/.

Jacoby B., 2009. Regression III Advanced Methods: Michigan State University, Lecture 14 Last accessed 20 November 2014. Available at: http://www.learningace.com/doc/202322/afa8a7dfb86c08c21d4aa91eff25197b/lecture14

Jones M.T., 2005. Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk. *IMF Working Paper WP/05/219*

King J.E., 2002. Statistical Significance vs. Practical Significance: *Implications for Analysing CME Data*. Last accessed 12 June 2014. Available at: www.ccitonline.org/jking/homepage/Signif_poster.doc

Mays E.,2001. Handbook of Credit Scoring. *Library Edition*

Menard S., 2002. Applied Logistic Regression Analysis*, 106.*

Merriam-Webster Dictionaries. Copyright© by Merriam-Webster, Incorporated. Last accessed 30 March 2015 Available at: http://www.merriam-webster.com/dictionary/effect

Microsoft. (2010). Microsoft Excel [computer software]. Redmond, Washington: Microsoft.

Middleton M., 2007. Introduction to Decision Trees. *Chapter 14*

Moody’s Investor Service., 2009. Moody’s Rating Symbols and Definitions. Available at: https://www.moodys.com/sites/products/AboutMoodysRatingsAttachments

Mordock J.B., 2000. Outcome assessment: Suggestions for agency practice. *Child Welfare League of America Journal*, 57.

Mulder A., 2008. Large Corporate Statistical PD Model. *Version 2.2.*

Nagelkerke N.J.D., 1991. A note on the general definition of the coefficient of determination. *Biometrika*, *78*(3).

Neter J., Kutner M.H., Nachsheim C.J. and Wasserman W. 1996. Applied Linear Statistical Models. *4th Edition.*

Osteen, P., & Bright, C. L., 2010. Effect sizes and intervention research. Workshop presented at the annual meeting of the Society for Social Work and Research, San Francisco, CA.

Pampel F.C., 2000. Logistic Regression: A Primer. *Sage University Paper on Quantitative Applications in the Social Sciences, 07-132.*

Parzen M. and Lipsitz S.R., 1999. A global goodness-of-fit statistic for Cox regression models. *Biometrics*, *55.*

Peng C-Y.J., Lee K. L. and Ingersoll G., 2002. An Introduction to Logistic Regression Analysis and Reporting, Indiana University. Last accessed 28 May 2014. Available at: https://datajobs.com/data-science-repo/Logistic-Regression-[Peng-et-al].pdf

Pindyck R.S. and Rubinfeld D.L., 1991. Econometric Models and Economic Forecasts. *3rd Edition*

Rice M.E. and Harris G.T., 2005 Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r. *Law and Human Behaviour. 29*(5).

Rodriguez G., 2001. Review of Likelihood Theory. *Appendix A.* Available at http://data.princeton.edu/wws509/notes/a1.pdf

SAS Institute Inc. 2011. The SAS System for Windows Enterprise Guide version 5.1 Copyright© by SAS Institute Inc., Cary, NC, USA. http://www.sas.com/en_us/home.html

SAS Usage Note 39724. 2011. ROC analysis using validation data and cross-validation. Last accessed 13 August 2014. Available at: http://support.sas.com/kb/39/724.html

Satchell S. and Xia W. 2006. Analytical Models of the ROC Curve: Applications to Credit Rating Model Validation. *Research Paper 181*

Scott Long J., 1997. Regression models for categorical and limited dependent variables. *1st Edition.*

Siddiqi N., 2006. Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring. *1st Edition.*

Sobehart J., Keenan S. and Stein R., 2000. Validation Methodologies for Default Risk Models. *5.*

SPSS Inc. (2014). IBM SPSS Statistics Version 22, Release 22.0.0,Copyright© IBM Corporation and its licensors. http://www-01.ibm.com/software/analytics/spss/

StatSoft, Inc. (2014). STATISTICA (data analysis software system), version 12. http://www.statsoft.com.

Steyn H.S., 2009. Manual for the determination of Effect size indices and Practical Significance. Last accessed 14 July 2014. Available at: http://www.nwu.ac.za/content/statcs-manual

Swanson E., Saviano C. and Zha L., 2010 Akaike Information Criterion versus Schwarz Bayes Information Criteria, Final Project: Reed College

Whitehead J., 2011. An Introduction to Logistic Regression, Appalachian State University. Last accessed 28 May 2014. Available at: http://www.appstate.edu/~whiteheadjc/service/logit/intro.htm#logitmodel

Wilkinson and the American Psychological Associates (APA) Task Force on Statistical Inference. 1999. Statistical methods in psychology journals: Guidelines and expectations. *American Psychologist*, *54*(8).

Wilkinson and the APA Task Force on Statistical Inference. 2001. Publication Manual of the American Psychological Association. *5th Edition*